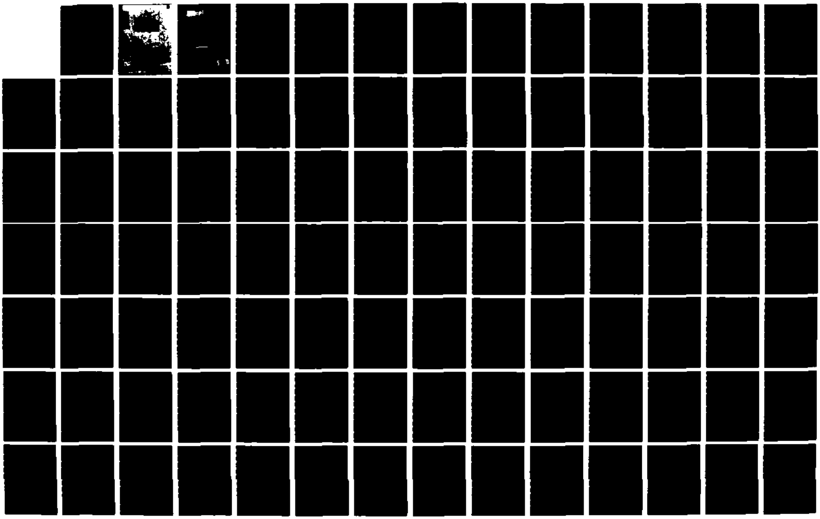
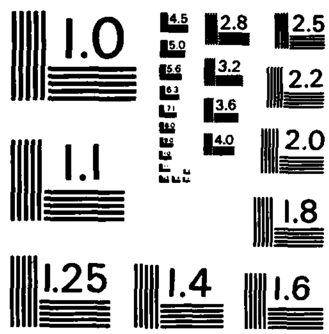
AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 1/9

UNCLASSIFIED 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

The Event-Related Brain Potential as an Index of Information Processing, Cognitive Activity, and Skill Acquisition

AFOSR

Prepared by:
The Cognitive Psychophysiology Laboratory

Prepared for:
The Air Force Office of Scientific Research
Life Sciences Directorate

2

cpl

COGNITIVE
PSYCHOPHYSIOLOGY
LABORATORY

Department of Psychology
University of Illinois
Champaign, Illinois 61820

Technical Report No. CPL 85-1
F49620-83-C-0144
January 1985

The Event Related Brain Potential as an Index of Information Processing, Cognitive Activity, and Skill Acquisition:

A Program of Basic Research

Final
[redacted] Report

Prepared for:

The Air Force Office of Scientific Research
Life Sciences Directorate

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS
A154118

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
CPL 85-1

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 85-0662

6a. NAME OF PERFORMING ORGANIZATION
Department of Psychology
University of Illinois

6b. OFFICE SYMBOL *(If applicable)*
NL

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NL

6c. ADDRESS *(City, State and ZIP Code)*
603 East Daniel St.
Champaign, IL 61820

7b. ADDRESS *(City, State and ZIP Code)*
Bolling AFB DC 20332

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
Air Force Office of Scientific Research

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F49620-83-C-0144

8c. ADDRESS *(City, State and ZIP Code)*
Bolling Air Force Base, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313	A4	

11. TITLE *(Include Security Classification)* The Event-Related Brain Potential as an Index of Information Processing, Cognitive Activity, and Skill Acquisition

12. PERSONAL AUTHOR(S) Emanuel Donchin, Christopher Wickens, Michael G.H. Coles

13a. TYPE OF REPORT Final Technical

13b. TIME COVERED FROM 83Sep01 TO 84Aug31

14. DATE OF REPORT *(Yr., Mo., Day)* 85Feb28

15. PAGE COUNT 21pgs + 2 Appendices

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Event-related brain potential (ERP), P300, Memory, Mental Chronometry, Information Processing, Workload, Automatic versus Controlled Processing, Resource Reciprocity, Dual Tasks, Vector Filters.

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

We review a program of research designed to understand the event-related brain potential (ERP) so that it can be used as a tool in the study of cognitive function and in the assessment of man-machine systems. We have conducted a series of studies on the functional significance of ERPs and have demonstrated that the P300 component is related to memory processes. We have used measures of the same component to evaluate workload, to time mental processes, to study the reciprocity of processing resources, and to extend theories of human information processing. We have also made technical advances in the analysis of the distribution of electrical potentials across the scalp.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Alfred R. Fregly

22b. TELEPHONE NUMBER *(Include Area Code)*
202/767-5024

22c. OFFICE SYMBOL
AFOSR/NL

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

TABLE OF CONTENTS

	Page
1. The Research Strategy	1
1.2 The Event-Related Brain Potential (ERP)	1
1.3 The Psychophysiological Paradigm	2
2. Progress Report: The Last Project Year	3
2.1 Studies of Working Memory	4
2.1.1 A von Restorff Memory Experiment	4
2.1.2 Incidental Free Recall	7
2.1.3 Manipulation of Memorization Strategies	7
2.1.4 Sternberg Experiment	8
2.2 Mental Chronometry	10
2.2.1 P300 and Error Detection	10
2.2.2 Serial Stage Versus Continuous Flow Models	13
2.3 Applications of P300 to the Assessment of Workload	14
2.3.1 The Use of P300 in Task Analysis	14
2.3.2 Performance Enhancements under Dual Task Conditions	17
2.3.3 Automatic versus Controlled Processing	18
2.4 Technical and Methodological Advances	19
2.4.1 Vector Filters	19
2.4.2 The Consistency of ERPs	20
3. Appendix A: List of Publications and Papers for Project Period	
4. Appendix B: Selected Publications	

1. The Research Strategy

Note: Numbers in the text refer to the reference list given in Appendix A. Where the article, paper, or chapter is included in Appendix B, this is indicated by "B#".

Cognitive Psychophysiology, as its name implies, is a marriage of cognitive psychology and psychophysiology. The basic premise of this union is that the understanding of cognitive processes can be enhanced by augmenting the traditional tools of the cognitive psychologist by adding tools based on the measurement of physiological functions (1,3,4,6,7,10,24; see also B1, B9). The psychophysiological data are, of course, useful only to the extent that they complement and expand the view of the mind that can be developed with the use of more traditional techniques. This is the premise that underlies the research described in this report.

1.2 The Event-Related Brain Potential (ERP)

The ERP is a series of voltage oscillations that are time-locked to an event. It is derived by averaging samples (epochs) of the electroencephalogram (EEG) recorded from the human scalp with each sample having the same temporal relationship to a particular event. Note that we can look at activity preceding an event, as well as activity following an event. The voltage oscillations derived in this manner are regarded as manifestations of different "components". Components are defined in terms of their polarity (positive or negative voltage), latency range (temporal relationship to the event), and scalp distribution (variation in voltage with electrode location on the scalp), as well as by their relationship to experimental variables. Components can be quantified using simple magnitude measures or through the application of more advanced techniques such as

Principal Component Analysis (PCA) and Vector Analysis (2,12,13,27; see also B2, B5, & B11). They are labeled by a polarity descriptor (P or N for positive or negative) and a modal latency descriptor (e.g. 300, for 300 msec). Thus, the P300 is a positive ERP component with a modal latency of 300 msec. In some cases, as with Contingent Negative Variation (CNV) and Slow Wave (SW), the descriptors are omitted.

1.3 The Psychophysiological Paradigm

We assume that the voltages we record at the scalp are the result of synchronous activation of neuronal ensembles whose geometry allows their individual fields to summate to a field whose strength can affect scalp electrodes. It is convenient to parse the ERP into a set of components. The component, in our scheme of things, is characterized by a consistent response to experimental manipulations. We further assume that each component is a manifestation at the scalp of an intracranial processing entity. We are not implying that each ERP component corresponds to a specific neuroanatomical entity or that the activity manifested by the component corresponds to a distinct neural process. Rather, we assume that a consistent information processing need, characterized by its eliciting conditions, activates a collection of processes that, for perhaps entirely fortuitous reasons, have the biophysical properties that generate the scalp-recorded activity.

As a working hypothesis we postulate that ERP components are manifestations of functional processing entities that play distinct roles in the algorithmic structure of the information processing system. In other words, we believe that it is possible to describe in detail the transformations that the processing entity applies to the information stream. The

goal of Cognitive Psychophysiology, within this framework, is to provide such detailed descriptions. This may be achieved by developing comprehensive descriptions of the conditions governing the elicitation and attributes of the components (the "antecedent" conditions). These descriptions can be used to support theories that attribute certain functions to the subroutine manifested by the component. In turn, the theories should lead to predictions regarding the consequences of the elicitation of the subroutines, predictions that can be tested empirically.

2. Progress Report: The Last Project Year

This section reviews work conducted at the Cognitive Psychophysiology Laboratory (CPL) during the period 10/1/83-9/30/84, with the support of the present project.

In the main, the CPL continued in this period to pursue closely related goals. The primary mission of our research is to develop an understanding of the Event Related Brain Potential (ERP) so that it can be used as a tool in the study of cognitive function and in the assessment of man-machine interactions. To this end, we have conducted studies that fell into four, not altogether distinct, categories, as follows:

A. The elucidation of the functional significance of the ERPs in relation to memory

B. The use of ERPs in studies of cognitive workload

C. The use of ERPs in studies of mental chronometry

D. Methodological studies

Below, we present a systematic review of this research. A list of publications and presentations given during the project period is shown in

Appendix A. Appendix B (1 through 14) contains a selection of articles, chapters and abstracts.

2.1 Studies of Working Memory

In our studies on the functional significance of P300, we have focussed on the relationship between P300 and memory. As our understanding of the functional significance of P300 develops, and is clarified, our ability to use P300 in studies of human information processing will increase. With a comprehensive theory of P300 we will be able to utilize both P300 latency and amplitude as tools in the study of cognition.

To elucidate the functional significance of P300 we have been studying the relationship between memory and P300 amplitude. This work derives from the hypothesis that the P300 is elicited during "context updating". We postulate that the updating of representations in working memory is accompanied by a P300, and that the amplitude of the P300 is, in some sense related to the magnitude or strength of this updating process. This hypothesis leads to the prediction that subjects would be more likely to recall events that had initially elicited a large P300 than events which elicited smaller P300s. We will briefly review three experiments, which were designed to test this prediction, and a fourth experiment that deals with a different aspect of memory.

2.1.1 A von Restorff Memory Experiment (14)

We used a von Restorff paradigm to study the relationship between the P300 elicited when a word was presented and the subsequent recall of that word. In the von Restorff paradigm the subject is instructed to recall a series of items, and a deviant item (an "isolate") is embedded in the

series. Von Restorff, the Gestalt psychologist who created this paradigm, demonstrated that the isolates were better recalled by subjects than were comparable non-deviant items. This enhanced recall of the isolates is the von Restorff, or isolation, effect. As the isolates are both rare and task relevant, they are apt to elicit large P300s. Furthermore, we know that whether or not it was subsequently recalled. We ran 12 female subjects, who were shown series of 15 words. No word was ever repeated. The isolates were displayed with a larger, or smaller font, than the rest of the series, and the ERP's associated with each word were recorded. At the end of each list, the subject wrote all the words that she could remember. Our main experimental hypothesis was that isolated words recalled in the subsequent test should elicit larger amplitude P300s when initially presented than isolated words later not recalled. In order to test this hyothesis the ERPs recorded at the initial presentation of each word were sorted according to the outcome of the subsequent recall test.

The von Restorff index (VRI), a measure of the magnitude of the von Restorff effect, and an overall performance index (P) for the recall test were computed for each subject. The VRI indicates how much the recall of the isolates is enhanced with respect to the recall of the non-isolates. We found striking individual differences in the degree to which subjects showed the von Restorff effect. These differences were surprising, because this effect has always been described as very robust. We divided the subjects into three groups, according to the magnitude of their von Restorff effect. The two extreme groups are very different. Subjects in group 1 have a high VRI, that is they show enhanced recall for the isolates. They are poor memorizers. These subjects report using rote strategies (they merely repeat the words). For these subjects, isolates that were recalled elicit larger

amplitude P300s than isolates that were not recalled. On the other hand, subjects in group 3 do not show any VRI and are very good memorizers. They report using elaborative startegies (making up stories, or combining words into images, sentences, etc.) to memorize the words. For these subjects P300 amplitude is not related to recall. The amplitude of a frontal-positive slow wave was correlated with recall in these subjects.

These data support our position that P300 amplitude is the manifestation of an updating process in working memory. The data confirm that representation of a word in working memory is affected, in some manner, when a P300 is elicited. The change in the representation aids recall in rote memorizers. All subjets produced equally large P300s, and we believe the same updating process occurred in all our subjects. If no further processing occurs, as in our group of rote memorizers, then P300 amplitude will be related to recall. However, if cognitive activity continues after the initial processing reflected by P300, then this additional activity may obscure the relationship between P300 and recall. When subjects link words together the recall of any individual word becomes less dependent on its initial encoding, and more dependent on its relationships to other words. Thus, for example, a word that initially elicits a tiny P300 may still be recalled if it is linked to other words that are recalled.

Two experiments have been undertaken after this study was concluded. They are designed to test the reliability and generality of the phenomenon we discovered and to test the interpretation we proposed for the relation between recall and P300 amplitude.

2.1.2 Incidental Free Recall (11)

Subjects were exposed to series of visually-presented male and female names, and were required to count one class of names. Afterwards, subjects were unexpectedly asked to write down as many names (both male and female) as possible from the previous list. All subjects expressed surprise, as there was no indication that such a recall test would occur. In this experiment we did not expect large individual differences because the only processing required during the presentation of the names was to keep a mental count of the number of names of one gender. We expected a strong relationship between the P300 elicited by names when they were initially presented and later recall. The results confirmed this hypothesis. There was a statistically significant difference between the P300 amplitude elicited by names recalled vs. names not recalled. Names that were recalled elicited larger P300s during the initial presentation. An interesting aspect of the data was the appearance of what seems to be a secondary P300, with a latency of 900 msec. This P300 was particularly prominent in the words that were subsequently recalled. It is not yet clear how to interpret this component, and we are continuing our analysis, but in some cases it appears that two P300s may have been generated, while in other cases there is a very late P300, or a continued positivity after the peak P300 amplitude is reached.

2.1.3 Manipulation of Memorization Strategies (26; see B10)

A straightforward test of the suggestion that the P300/recall relationship is contingent on the subjects' recall strategies is a demonstration that the same subject can show both patterns when both strategies are employed. We have, therefore, manipulated subjects'

strategies to see if the pattern of results that we observed in different subjects can be reproduced within the same subject operating under different instructions. We used the von Restorff paradigm described before. In two experimental sessions, we gave 12 subjects explicit instructions about the strategies to use in memorizing the words. Preliminary analysis of the data suggests the following conclusions. First, subjects will change their strategies following instructions. Second, when they use the rote strategy, the von Restorff effect is large, overall recall is low, and P300 is related to later recall. Conversely, when the same subject uses elaborative strategies, the von Restorff effect is small, overall performance is high, and there is no relationship between P300 and later recall. These data are, of course, preliminary. However, they do suggest that P300 is related to a particular kind of memorial process.

2.1.4 Sternberg Experiment (30; see B13)

Our second approach to the analysis of ERPs and memory has involved the use of the paradigm developed by Sternberg to analyze the timing of search through memory. In these studies we utilize P300 latency, as well as its amplitude, as a dependent variable. The study illustrates the manner in which P300 can be used as an index response in cognitive psychology.

Forty-five subjects were presented with memory sets ranging from one to five letters. Thirty probes were then presented, one every two seconds, and subjects were to determine if the probe matched one of the elements in the memory set. Subjects were instructed to respond by pressing one of two buttons as rapidly as possible without making errors.

The reaction times followed the pattern reported by Sternberg. RT increased linearly as a function of set size for positive and negative

APPENDIX A

Publications and Papers for Project Period

APPENDIX #	CHAPTERS
B1	1. Coles, M. G. H., & Gratton, G. Psychophysiology and contemporary models of human information processing. In D. Papakostopoulos & I. Martin (Eds.), Clinical and Experimental Neuropsychophysiology. Beckenham, England: Croom Helm, in press.
B2	2. Coles, M.G.H., Gratton, G., Kramer, A., & Miller, G. A. Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press, in press.
	3. Donchin, E. The dissociation of electrophysiology and behavior: A disaster or a challenge? In E. Donchin (Ed.), Cognitive Psychophysiology. Hillsdale, NJ: Erlbaum Associates, 1984.
	4. Donchin E., Coles, M. G. H., & Gratton, G. Cognitive psychophysiology and preparatory processes: A case study. In S. Kornblum & J. Requin (Eds.), Preparatory States and Processes. Hillsdale, NJ: Erlbaum Associates, 1984.
	5. Donchin, E., Gratton, G., Dupree, D., & Coles, M. G. H. After a rash action: Latency and amplitude of the P300 following fast guesses. To appear in M. Kietzman, G. Galbraith, & E. Donchin (Eds.), Neurophysiology and Psychophysiology: Basic Mechanisms and Clinical Applications. New York: Academic Press, in preparation.
	6. Donchin, E., Karis, D., Bashore, T., Coles, M. G. H., & Gratton, G. Cognitive psychophysiology and human information processing. In Coles, M. G. H., Donchin, E., & Porges, S. (Eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press, in press.
	7. Donchin, E., Kramer, A., & Wickens, C. Applications of event-related brain potentials to problems in engineering psychology. In Coles, M., Donchin, E., & Porges, S. (Eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press, in press.
	8. Fabiani, M., Gratton, G., Karis, D., & Donchin, E. The definition, identification and reliability of measurement of the P300 component. In P. Ackles, J.R. Jennings, and M.G.H. Coles (Eds.), Advances in Psychophysiology, Vol. 3, Greenwich, CT: JAI Press, in press.

occurred and a different button with the other thumb when the other pitch occurred. The tones differed in probability (20% and 80%). The rare tone typically elicits a larger P300 than the frequent tone. Task 3 was somewhat different. Only one tone pitch was used. On 10% of the trials, the tone was not presented. The subject was to count the number of "omitted stimulus" trials. P300 is usually larger on such trials. Task 4 was a visual analog of task 2. Male names were presented on 20% of the trials, female names on 80%. Again, the rare class of stimuli should elicit a larger P300.

We (8) assessed the reliability of different measures of P300 amplitude and latency. Reliability was measured both at a single recording session ("within-session reliability"), and over time across two recording sessions ("between-sessions reliability"). Several measurement procedures were compared, including peak, area, cross-correlation, and PCA measures. Measures obtained at Pz were compared with measures obtained with the Vector filter procedure (2,12,13; see B5). The reliability of P300 measures was generally high (up to .92 for amplitude measures and .83 for latency) and depended on the number of trials and on P300 amplitude. The highest reliability for amplitude measures was obtained with a cross-covariance measure combined with a vector filter procedure applied on single trials. The highest reliability for latency measures was obtained with peak-picking combined with vector filter. In general, averages of single trial estimates were more reliable than measures taken directly on averages waveforms. Between-session reliabilities were lower than within-session reliabilities, but still usually higher than .60.

We have developed a method which permits the assessment of the degree of similarity between an obtained ERP distribution and a distribution defined, a priori (2,12,13; see B5). Thus, for a single ERP trial, or for an average ERP, we can measure the "P300ness" of each point in the waveform. This procedure can be conceptualized as filtering the ERP for its distributional characteristics. This "vector filter" procedure permits an asessement of both P300 amplitude (the maximum value of the filter output) and latency (the timepoint of this maximum) for both single trials and average ERPs.

We have recently completed a series of simulation studies in which we compare a vector filter analysis of the latency of the P300 with traditional techniques (see B5). The latency estimates using vector filter are both more reliable and valid than those based on other algorithms.

2.4.2 The Consistency of ERPs (8)

We have begun to evaluate the consistency of various aspects of the ERP since a basic issue in the application of ERPs in the assessment of human operators is the consistency across situations of the ERP generated by a given operator. The more consistent an individual's waveform across tasks, the more reliably his or her performance can be monitored under changing circumstances. To date, we have run a sample of 20 young adults in four tasks which were chosen to produce P300s which then could be evaluated for consistency across the four tasks.

Task 1 required subjects to count the number of occurrences of one of two equiprobable tones which differed in pitch. In general, P300 is larger for the counted than for the uncounted tones in this paradigm. In task 2, the subject pressed a button with the left thumb when one tone pitch

condition following extensive practice. P300 latency mirrored RT, suggesting that the development of automatic processing substantially reduced stimulus evaluation time. The commonly observed relation between probability and P300 amplitude, with larger P300s elicited by infrequent events, was found in the VM conditions but not in the CM condition after practice. This suggests an attenuation of memory updating during automatic processing. Two different negative components were affected by stimulus mismatch. These components appear to reflect different degrees of mismatch processing.

2.4 Technical and Methodological Advances

We have continued to pursue our interest in methodological and technical advances which aid in the quantfication and analysis of ERPs. In this regard, we have also written a major methodological chapter for a new Handbook of Psychophysiology (2; see B2).

2.4.1 Vector Filters

Following our solution of the eye-movement artifact problem (see last year's report), we have turned our attention to another problem in the analysis of ERPs. This problem concerns the quantification of a component of the ERP when the definition of the component includes a distributional aspect. For example, the P300 is defined, not only in terms of its polarity and latency, but also in terms of its distribution across different scalp locations. It is seen most positively at the parietal electrode and least positively at the frontal electrode. The critical question is - how do you quantify distributional information?

effect demonstrated for P300 served as the basis for the analysis of the resource tradeoffs between dual-task combinations.

Twelve subjects participated in three experimental sessions in which they performed both single and dual-tasks. The primary task was a pursuit step tracking task. The secondary tasks required the discrimination between different intensities or different spatial positions of a stimulus.

Task pairs which required the processing of different attributes of the same object resulted in better performance than task pairs which required the processing of different objects. Furthermore, these same object pairs led to a positive relation between primary task difficulty and the resources allocated to secondary task stimuli. Inter-task redundancy, the physical proximity of task related stimuli and processing priorities also affected the performance of dual-task pairs. The results of the study have lead to a P300 based model of the conditions which influence the degree of integrality between dual-tasks.

2.3.3 Automatic versus Controlled Processing (17; see B6)

This study focused on the effects of, and the interactions between, practice and task structure on human performance. The development of automatic processing through consistent stimulus-response mapping (CM) was assessed by means of measures of reaction time and event-related brain potentials. The subjects performed a visual searcn task in which they responded by pressing a button whenever a probe matched a memory set item. The variables manipulated in the study included the number of memory set items (1 or 4), the task structure (CM or VM), and the probability of occurrence of a memory set item (.2 or .8). Set size had a significant effect on RT in both CM and VM conditions prior to practice and in the VM

salient perceptual/central processing component. The results might also be useful in the design and evaluation of complex tracking tasks. If operators are required to perform a manual control task with a multidimensional system and/or with higher order system dynamics then concurrently performed tasks should be designed so as to minimize perceptual/central processing load. We see here, again, how the ERPs provide data that increase the theoretical depth with which one can draw conclusions about the human information processing system.

2.3.2 Performance Enhancements under Dual Task Conditions (19; see B7)

Most research on dual-task performance has been concerned with delineating the antecedent conditions which lead to dual-task decrements. Capacity models of attention which postulate a hypothetical resource structure underlying performance have been employed as predictive devices. These models predict that tasks which require different processing resources can be more successfully time shared than tasks which require common resources. We have recently suggested that dual-task decrements can be avoided even when the same resources are required by both tasks, by designing the tasks so that the processing demands can be integrated. The conditions under which such dual-task integrality can be fostered were assessed in a study in which we manipulated three factors likely to influence the integrality between tasks: the redundancy between elements in two concurrently performed tasks, the physical proximity of two visual tasks on a CRT, and whether the tasks required the processing of the same or different objects. The resource structure associated with these integrated dual-task pairs was inferred from changes in the amplitude of the P300 component of the Event-Related Brain Potential. The resource reciprocity

therefore, that the P300 elicited by the intensifications of the target and cursor, associated with an oddball task run concurrently with the tracking task, would be larger during the acquisition than during the alignment phase.

The P300 amplitude was attenuated both as a function of the phase of the tracking task, larger amplitude P300s elicited in the acquisition phase, and system order, larger P300s elicited during the easier, first order tracking. Thus, P300 amplitude has proven to be sensitive to changes in resource demands both within single trials (tracking phase) and across different experimental conditions (system order). This study, along with additive factors investigators of manual control parameters, have provided converging evidence that system order has a salient perceptual/central processing component.

Another study recently completed in the Cognitive Psychophysiology Laboratory has provided additional insight into the resource demands of tracking dimensionality (29; see B12). In this study subjects were required to perform a pursuit step tracking task as their primary task. The secondary task was an auditory oddball. The difficulty of the tracking task was manipulated by changing the system order (first or second) and the number of dimensions to be tracked (one or two). Consistent with previous research, P300 did not discriminate between the number of dimensions when the subjects were tracking with the first order system. However, P300s were systematically affected by the number of dimensions when the subjects performed with a second order system. The P300s elicited by the counted tones decreased in amplitude with increases in the number of tracking dimensions. This result suggests that multidimensional systems which are relatively difficult to control (e.g., with second order dynamics) possess a

dynamics requires a large degree of perceptual anticipation as well as a modified response strategy. Assuming that P300 amplitude is sensitive to the perceptual aspects of a task then a reduction in P300 amplitude by higher order control should localize some of the influence of the order variable at the earlier processing stages.

The subjects' primary task was as follows. A target appeared on the screen and moved in a straight line at a randomly selected angle. The subject had to move a cursor into the neighborhood of the target. The time between the appearance of the target and its acquisition by the cursor is called the "acquisition phase". Acquisition was accomplished by manipulating the two-axis joystick mounted on the right side of the subject's chair. Successful acquisition initiated the alignment phase. The target began to rotate at a constant velocity in either a clockwise or counterclockwise direction. The subjects had to rotate the cursor at the same velocity as the target while also keeping the two elements superimposed. The rotation was accomplished by manipulating a single axis joystick mounted on the left side of the subject's chair. A deflection of the stick to the right produced a clockwise rotation of the cursor at an angular velocity proportional to the angle of deflection, a deflection to the left produced a counterclockwise rotation. Deviation from the initial acquisition criterion for more than 1000 msec necessitated a re-alignment of the elements. Once the subjects decided that all of the criteria had been satisfied and the target and cursor were aligned, they could press a capture button and the trial was terminated.

We assumed that the alignment phase would be more difficult than the acquisition phase due to increased perceptual demands imposed by the requirement to control the additional rotational axis. We predicted,

evaluation system to a response activation system before the evaluation process is completed. In this sense, our data are more consistent with continuous flow models of information processing than with serial stage models.

2.3 Applications of P300 to the Assessment of Workload

Several studies have been completed and/or published in the last year which have provided additional insight into the information processing of subjects during the performance of complex manual control tasks. These studies have focused on using P300 to decompose the processing requirements of higher order control systems, investigated the resource requirements of tracking dimensionality at different levels of system order, mapped the resource tradeoffs of dual-task pairs, and explicated the factors which result in enhancements in dual-task performance. We have also written a chapter on the measurement of workload (see B3).

2.3.1 The Use of P300 in Task Analysis

Wickens, Kramer, and Donchin (30; see B14) performed a componential analysis of the demands of controlling higher order systems. By "order of control" we refer to the number of time integrations between the output of a controller (i.e., joystick) and the output of the system. In a first order, or velocity driven system, a deflection of the joystick corresponds to a change in the velocity of the controlled element. A second order, or acceleration driven system, produces a change in the accleration of the controlled element proportional to the deflection of the control stick. The increase in system order appears to increase the demand for both perceptual and response related resources. Effective control over second order

According to this view the elicitation of the P300 is delayed on the error trials because the system is aware of the error and engages in additional processing before the trial information can be accomodated in the subject's world model.

2.2.2 Serial Stage Versus Continuous Flow Models (1,10; see B1 & B4)

We conducted an experiment that was designed, in part, to use psychophysiological measures to evaluate different models of human information processing. In this experiment, we used the measure of P300 latency to assess the time it takes a subject to evaluate a stimulus (22; see B8). We also used measures of the electromyogram and "sub-threshold" behavioral responses to define different types of trials in terms of the degree of error present. Specifically, in a choice reaction time task, we find that subjects sometimes initiate responses with the incorrect hand, although the complete response is actually made with the correct hand. These trials may be thought of as "partial" error trials. Subjects were required to make a discriminative response to the center letter in a five letter stimulus array. For some arrays, the noise letters surrounding the center letter were the same as the center letter; for other, incompatible arrays, the noise letters were those associated with the opposite response. We find that there are more error and partial error trials for incompatible arrays. These errors and partial errors lead to a delay in the production of the correct response. Our data also show that as P300 latency increases, the probability of error increases, and that for a given P300 latency, the probability of error is greatest if the subject responds quickly. If we assume that P300 latency is a measure of stimulus evaluation time, then these data support the notion that information is passed from a stimulus

- The reaction times associated with these error trials were in general very fast. Correct responses to Male names were considerably slower.
- The reaction times to Female names were in general as fast as were the reaction times to Male names. Though, in both cases there was a similar distribution of reaction times.

It seems from the above, and from analyses that we do not have the space to describe in this report, that the subjects' behavior suggests that in both the Speed and the Accuracy conditions a bias to press the Female button was maintained. Subjects' responses were thus driven largely by this bias. Alternate models were tested and were not consistent with all aspects of the data set.

The ERP data can be summarized as follows:

- The Male names in all series elicited a substantial P300, characterized by the scalp distribution commonly observed for the P300.
- Female names elicited a very small and indistinct P300 when the probability of such names was .80.
- The latency of the P300 elicited by Male names was considerably longer when the subject erred on the trial than it was when the subject was correct. That is, for those male names that were responded to slowly, and correctly, the P300 latency was shorter than it was on those trials in which the subject responded very fast.
- Female names that were responded to with equal speed as were the error triggering male names did not elicit a delayed P300. In other words, it is unlikely that the longer P300 on error trials is due merely to the fast responses made on these trials.

respond by pressing the "female" button. Again, the results suggested that for all subjects, the P300 latency was increased on these error trials. There remained, however, a number of questions. It was not possible, for example, to determine if the increased latency was due to the fact that an error was committed or to the fact that the response tended to be fast on these trials. It was also not possible to determine from these data the extent to which the emphasis on speed was critical for the pattern of results. Some investigators doubted that the component we identified was a delayed P300. It was suggested that the delayed peak represents a new component rather than a delayed P300.

We decided, therefore, to conduct an investigation that would try, in the design of the experiment, to address most of these concerns. To this effect we have run 7 male subjects, each in 4 conditions obtained by combining two levels of probability (p[male] = .50 and .20) and two instruction regimes (speed and accuracy). Data were recorded on 800 trials in each of the 4 cells from each of the four subjects.

The data on the subjects' overt responses could be summarized as follows:

- The subjects appeared to have adopted the instructional regimes as they tended to respond faster when instructed to be fast. Reaction times were longer, and the errors fewer when accuracy was emphasized.
- In the speed conditions, the subjects hardly ever pressed the Male button in response to a Female name. However, they made substantial Female button presses in response to Male names.

negative probes, the match will be slower since the item to be matched is lower down in the stack. Note that, in the case of the positive probes, the processes associated with RT and P300 are coupled - hence, the RT/P300 latency correlation. For negative probes, however, RT and P300 are related to quite different processes - RT depends on the failure to find a match within the positive item set, while P300 depends on the presence of a match with an negative item. Hence, the decoupling of RT and P300.

2.2 Mental Chronometry

2.2.1 P300 and Error Detection (5)

We have explored the functional significance of the ERP under circumstances in which the subject makes an error in responding to a stimulus. A tantalizing observation that recurred in many of our studies in mental chronometry has been that on trials on which the subjects appear to be responding hastily, the P300 latency tends to be unusually long. This pattern appeared first in the study reported by Kutas, McCarthy, and Donchin. The subjects were instructed to count the number of times names of males appeared in a list of common names. Some 80% of the names on the list were names usually ascribed to females. When the subjects were urged to be as fast as possible they tended to respond with a very short reaction time on the "female" button, even when the name presented was a "male" name. Strikingly, all these fast guesses were associated with long P300 latencies.

The conditions of the first study did not provide for the occurence of a large enough number of these trials to allow for very firm conclusions. McCarthy, Kutas and Donchin replicated the study using a much larger number of trials and urging the subjects even more to be fast. Indeed the number of errors increased greatly. The subjects appeared to be very biased to

probes. Negative probes were associated with longer reaction times than positive probes. The slope of the regression lines for positive and negative probes were essentially the same. The standard deviation of RTs increased as a function of set size for both positive and negative probes. Error rates for all conditions were under 5%. These results are consistent with the findings reported by Sternberg, i.e., an exhaustive search process.

The ERP data have revealed that larger P300 amplitudes, and shorter P300 latencies are associated with positive rather than negative probes. P300 latency increases as a function of set size for positive but not negative probes. Note that this latter finding represents a clear dissociation between RT and P300 latency. For positive probes, both P300 latency and RT increase with set size. Furthermore, on a within-subject basis, P300 and RT are modestly but significantly correlated. For negative probes, however, RT, but not P300 latency, increases with set size. And, on a within-subjects basis, RT and P300 latency are not significantly related.

We interpret these data in the following way. The subject holds the items to be remembered in a "memory stack" - other letters of the alphabet may also reside in the stack but they are at a lower level than the positive items. For both positive and negative probes, the production of a response depends on a search through the positive items at the top of the stack. If a match between a probe and an item is made, the subject responds "yes" - if no match is made, the subject responds "no". In both cases, the subject apparently searches through all positive items before a response is made. This accounts for the reaction time data. The P300, on the other hand, is dependent on a different process - namely, the matching of a probe with an item in the stack. For positive probes, this match will be made faster and with more certainty, since the item is near the top of the stack. For

B3 9. Gopher, D. & Donchin, E. Workload - An examination of the concept. In K. Boff & L. Kaufman (Eds.), Handbook of Perception and Human Performance. New York: Wiley & Sons, in press.

ARTICLES

B4 10. Coles, M. G. H., Gratton, G., Bashore, T.R., Eriksen, C.W., & Donchin, E. A psychophysiological investigation of the continuous flow model of human information processing. Submitted to Journal of Experimental Psychology: Human Perception and Performance.

11. Fabiani, M., Karis, D., & Donchin, E. "Gee...That was my aunt Mathilda": P300 and recall in an incidental memory paradigm, in preparation.

12. Gratton, G., Coles, M.G.H., & Donchin, E. Vector analysis of event-related brain potentials, in preparation.

B5 13. Gratton, G., Kramer, A.F., & Coles, M.G.H. A simulation study of latency measures of components of event-related potentials, in preparation.

14. Karis, D., Fabiani, M., & Donchin, E. P300 and memory: Individual differences in the von Restorff effect, Cognitive Psychology, 1984, 16, 177-216.

15. Klein, M., Coles, M. G. H., & Donchin, E. People with absolute pitch process tones without producing a P300. Science, 1984, 223, 1306-1308.

16. Kramer, A.F. The interpretation of the component structure of event-related brain potentials: An analysis of expert judgments. Psychophysiology, in press.

B6 17. Kramer, A.F., Fisk, A., Schneider, W., & Donchin, E. An event-related potentials analysis of automatic and controlled processing, submitted to Psychophysiology.

18. Kramer, A.F., Ross, W., Hulin, C., & Donchin, E. The subjective dimensions of event-related brain potentials: A multidimensional scaling analysis, in preparation.

B7 19. Kramer, A.F., Wickens, C., & Donchin, E. The processing of stimulus properties: Evidence for dual-task integrality. Journal of Experimental Psychology: Human Perception and Performance, 1985, in press.

20. Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 1984, 21, 171-186.

21. Wickens, C.D. & Kramer, A.F. Engineering Psychology. Annual Review of Psychology, 1985.

ABSTRACTS AND CONFERENCE PRESENTATIONS

B8 22. Coles, M. G. H. Flies in the ointment: The use of P300 in mental chronometry. Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England, 1984.

23. Coles, M. G. H. Psychophysiology and contemporary models of human information processing. Biological Psychology, in press.

B9 24. Donchin, E. The use of ERPs to monitor non-conscious mentation. Presented at the 20th Annual Conference on Manual Control, Ames, California, 1984.

25. Donchin, E., Kramer, A., Mane, A., Karis, D., & Heffley, E. Psychophysiological tools in engineering psychology. In Proceedings of Psychology in the Department of Defense, Ninth Symposium. United States Air Force Academy, Colorado Springs, Colorado, 1984.

B10 26. Fabiani, M., Karis, D., & Donchin, E. Effects of mnemonic strategy manipulation in a von Restorff paradigm. Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England, 1984.

B11 27. Gratton, G., Coles, M.G.H., & Donchin, E. Component identification with Vector Analysis. Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England, 1984.

28. Mane, A. M., Coles, M. G. H., Karis, D., Strayer, D., & Donchin, E. The design and use of subtasks in part training and their relationship to the whole task. Proceedings of the 20th Annual Conference on Manual Control, in press.

B12 29. Sirevaag, E., Kramer, A., Coles, M. G. H., & Donchin, E. P300 amplitude and resource allocation. Psychophysiology, 1984, 21, 598-599.

B13 30. Strayer, D. L., Karis, D., Coles, M. G. H., & Donchin, E. P300 latency and reaction time are dissociated in a Sternberg task. Psychophysiology, 1984, 21, 600.

B14 31. Wickens, C.D., Kramer, A.F., & Donchin, E. The event-related potential as an index of the processing demands of a complex target acquisition task. Brain and Information: Event-Related Potentials, Vol. 425, reprinted from the Annals of the New York Academy of Sciences, 1984, 295-299.

APPENDIX B1

Psychophysiology and Contemporary Models of Human Information Processing

Michael G. H. Coles & Gabriele Gratton

To appear in D. Papakostopoulos and I. Martin (Eds.), Clinical and Experimental Neuropsychophysiology. Beckenham, England: Croom Helm, in press.

Psychophysiology and Contemporary Models of Human Information Processing [1]

Michael G. H. Coles & Gabriele Gratton

Cognitive Psychophysiology Laboratory

University of Illinois, Champaign, Illinois 61820

Cognitive psychologists are interested in how a particular input (stimulus) to the human information processing system is translated into a particular output (response). They propose that different processing structures perform different transformations on the input such that, ultimately, an output is produced. The number of structures, and their function, varies among theories; however, in general, they include perceptual, central, and response structures (e.g. Wickens, 1980).

Traditionally, it has been proposed that the processes associated with the different structures occur sequentially--that is, the process associated with one structure must be completed before another process begins. In recent years, these discrete models have been challenged by those who propose that information can be transmitted from one structure to another before the process performed by the first is completed. Thus, continuous models imply that several processes can occur simultaneously (or in parallel) and that a given process can operate on the partial information provided by another process. (See Miller, 1982, for a discussion of the difference between discrete and continuous models).

The measurements taken by cognitive psychologists (reaction time-RT, percent correct, etc.) are seriously limited in terms of their ability to test continuous theories, principally because they represent a single output measure which is determined by many intervening processes, and their interactions. Thus, a particular experimental manipulation may affect not only the duration

Donders argued that the three types of reaction time tasks differed in terms of the number of stages or processes involved in their successful execution. Thus, the c-reaction time task adds a process of discrimination to the a-reaction time task, while the b-reaction time task adds an additional process of response selection. If RT is measured in each task, then by subtracting the RTs for various types of task it should be possible, according to Donders, to identify the time taken for discrimination and response selection processes respectively. Note that the subtractive method advocated by Donders presupposes that (a) the various stages of human information processing are arranged serially, (b) the duration of each stage is causally independent of the duration of the other stages, and (c) it is possible to use an experimental manipulation to add a stage (or process) without affecting other stages. The latter assumption has been referred to as the "postulate of pure insertion" (Ashby and Townsend, 1980). Given these assumptions, we can see that RT represents the sum of the durations of several component processes, and that it is possible to determine the duration of a stage "inserted" by a manipulation.

Sternberg's model. Sternberg (1969a) proposed a similar model ("stage model") to that advocated by Donders. Like Donders, Sternberg assumed that the human information processing system consists of a number of serially arranged stages and that RT represents the sum of the durations of each stage. However, the two theorists differ in the interpretation of the effect of an experimental manipulation. While Donders believed that a manipulation results in the insertion of a stage, Sternberg argues that it affects the duration of a particular stage, without affecting the duration of other stages. This is referred to as the postulate of "selective influence" (see Pieters, 1983).

Discussion of serial models. Serial models have a basic appeal because of their simplicity and elegance, and this is probably responsible in part for their dominance in the field for more than a century.

The postulates of pure insertion and selective influence, if supported, imply that the cognitive psychologist has the relatively simple task of using a variety of experimental manipulations to identify the number and function of the information processing stages. This task is accomplished using straightforward statistical procedures such as "t"-test or analysis of variance. Unfortunately, the real world is not so simple.

First, even a cursory knowledge of biological systems suggests that the nervous system does not function in a serial manner. "For," as Woodworth (1938) said, "there is nothing to prevent two cerebral processes from occurring simultaneously. Two responses, one perceptual and one motor, may take their start simultaneously from the same stimulus. If the motor response is not made to depend on the perception, it can start at once, as in the simple reaction. The brain is not a one track road " (p. 305).

Second, it has been argued (Pachella, 1974, p. 57) that the additive factors method relies on the acceptance of the null hypothesis for the inference of "independence". That is, if the interaction between two experimental factors is not significant, then it is claimed that the two factors influence different stages. This is clearly a risky procedure, unless great care is taken to insure sufficient statistical power.

Third, it has been argued that the postulates of pure insertion and selective influence have no basis in experimental data (Pachella, 1974). These postulates assume that the information processing sequence and the activities of the different stages remain unaffected by an experimental manipulation that either adds, or influences, a particular stage. The postulates can be

In this section we shall review the proposals of three of the principal adherents of the new approach, McClelland, Grice, and Eriksen. Rather than describe each of their views in detail, we shall focus on the special aspects of their proposals. We shall also consider the views of Miller who advances a theoretical framework that encompasses both serial and parallel approaches.

McClelland's cascade model. As its name implies, McClelland's model proposes that the human information processing system consists of a collection of processes arranged "in cascade" (McClelland, 1979). Although the processes are ordered, each process is continuously active. It operates on the basis of its input, and, at any particular time, provides an output that corresponds as closely as possible to an ideal transformation of the input. Since the transformation takes a finite time, each process introduces a delay into the system. An important assumption of the model is that any piece of information that enters the system must proceed in an orderly manner through all the processes. No side trips, back-tracking, or short-cuts are allowed. Furthermore, an important exception to the continuous activation of the processes is the response activation system. This system is assumed to operate in a discrete manner. When the input to this processing level exceeds a prescribed level, a response is emitted (see Grice, Nullmeyer, and Spiker, 1982; and see below).

McClelland provides a mathematical formulation to describe the behavior of each process. For present purposes, it is sufficient to note that the output of each process is assumed to be an exponential function of its input.

Grice's general theory. This theory was developed with the express intention of accounting for Donders' three types of reactions. Like McClelland, Grice believes that different aspects of the human information processing system can be active at the same time (Grice et al., 1982). Thus,

reaction time is prolonged. This process of response competition is a special feature of Eriksen's model.

Miller's "grain" hypothesis. The basis for Miller's hypothesis is the view that the distinction between discrete and continuous models of human information processing is an oversimplification (Miller, 1982). This distinction can be described in terms of the way in which information is transmitted. A critical concept for Miller is that of "grain". He uses it to refer to the units in which the information is transferred. For discrete models, information is transmitted in large "whole" grains--that is, all the information about the stimulus is passed on at one time. For continuous models, information is transmitted in infinitely small grains--that is, information is passed on continuously. Extending this logic, Miller argues that there may be circumstances where the grain is neither "whole" and large nor infinitely small. In this case, partial information may be transferred but the number of units is finite and may be small. He argues that these units may correspond to mental codes such as those for letters and numbers.

In a series of ingenious experiments in which partial information about the stimulus is given in advance of stimulus presentation, Miller has demonstrated that grain size varies with experimental conditions.

Conclusions

In this section, we have reviewed two classes of information processing models. While the propriety of the two classes may vary with experimental situation (Miller, 1982), several investigators believe that continuous models provide the more veridical description of the nature of human information processing in most situations.

outputs emitted at different levels of the information processing flow. Such a representation also illustrates the value of psychophysiological responses as "mapping devices." Variables which effect the flow of information above the psychophysiological response output will also affect that response. Those variables which affect the flow of information below the psychophysiological response output will not affect that response. In the same way, the time course of the processes may be reflected by parallel variations in the time course of psychophysiological responses. Furthermore, we can derive hypotheses about those processes which affect psychophysiological responses, but which do not affect current overt behavioral responses.

Given these considerations, several criteria are applicable to the selection of suitable measures of psychophysiological activity. First, the measures should have temporal properties (latency, etc.,) which are similar to those of the processes of interest. This criterion is particularly appropriate for studies of mental chronometry, where the interest is in the timing of mental processes. In the case of preparatory processes, we would expect there to be a similarity between the time courses of both the psychophysiological changes and the processes of which the changes are manifestations. Second, psychophysiological and traditional "behavioral" measures should be dissociable, at least under some circumstances (Donchin, 1982). If psychophysiological measures and behavioral measures are perfectly correlated, then the former will merely serve as "substitutes" for the latter. Such redundancy would trivialize the value of psychophysiological measures, since measures of reaction time are easier and cheaper to obtain. Third, the measure should be a manifestation of the activity of a structure involved in human information processing. In some cases, there may be a match between the process manifested by the psychophysiological measure and a process proposed by

CNV. The contingent negative variation (CNV) is a negative going component of the ERP. It is observed in the interval between two stimuli when some contingency has been established. Although there is some controversy concerning the functional significance of the component, it is generally believed that it is, at least in part, related to motor preparation (see Donchin, Coles, & Gratton, 1984). Thus, we propose that the CNV can be used as a marker for the presence of motor preparation--or the activation of response-related processes.

EMG. The electromyogram (EMG) is generated by the electrical activity of the muscles--and is therefore a manifestation of muscle activity. Our interest in this measure is focussed on the observation that (a) muscle activity preceeds movement, (b) EMG is sensitive to "subliminal" muscle activity that does not result in an "overt" response (movement).

An Example - H and S

We will now review an experiment in which we used the psychophysiological approach to evaluate Eriksen's continuous flow model (Coles, Gratton, Bashore, Eriksen, and Donchin, in preparation).

The particular setting we chose for a test of the approach was an apparently simple one. Twelve male subjects were required to make a discriminative response as a function of the center letter (target) in a five letter stimulus array. There were four arrays: HHHHH, SSSSS, HHSHH, and SSHSS. The responses we required of the subject were slightly unusual--a squeeze with the left or right hand of zero-displacement dynamometers at 25% of maximum force.

A critical aspect of the continuous flow model is that it proposes that the activation of the incorrect response interferes with the execution of the correct response thereby postponing reaction time. To analyze this process of response competition, we used measures of EMG activity and squeeze activity on the incorrect side to classify trials in terms of their degree of error.

N - Activity only on the correct side in EMG and squeeze channels

E - Activity on the correct side for EMG and squeeze channels: activity also present for EMG on the incorrect side.

S - Activity on the correct side for EMG and squeeze channels: activity also present for both EMG and squeeze channels on the incorrect side.

Error - Activity on the incorrect side for EMG and squeeze channels. EMG activity on the correct side may or may not be present.

We will first review evidence concerning the nature of the compatibility effect (see Figure 1).

Insert Figure 1 About Here

The upper part of Figure 1 illustrates that S and Error trials (where the wrong squeeze response was produced) occurred more often when the arrays were incompatible. The lower part of Figure 1 illustrates two main points. First, the latency of activity on the correct side increases as the degree of activity on the incorrect side increases--that is, correct responses were longer when there was activity on the incorrect side (E and S categories). Second, the

stimulus evaluation time, the more likely it is that the subject will activate the incorrrect response.

Our second question concerned the effect of the non-informative warning tone (non-informative in terms of response choice). Why does the warning stimulus speed up responses?

The relevant data are shown in Figure 2. First, note that for all

----

Insert Figure 2 About Here

----

response classes, both EMG and squeeze latencies are shorter for the warned condition. Second, note that P300 latency (that is, stimulus evaluation) is not afffected by the warning. Third, note that warned trials are associated with a slightly higher incidence of incorrect activity. This is most evident for the S category.

Thus, the effect of the warning stimulus is to decrease response latency by about 30 msec, and increase the incidence of incorrect activity (S) by about 3%. At the same time, the warning stimulus does not affect the latency of P300 (stimulus evaluation time). These findings are most parsimoniously interpreted in terms of the speed-accuracy trade-off function. The warning tone leads the subject to adopt a less conservative strategy.

We believe that this strategy is best considered in terms of an "aspecific activation" process--that is, activation of the response channels can occur independent of the specific nature of the stimulus. In the case of the warning, this activation occurs during the foreperiod and may be manifested by the large CNV that is present.

Variations in the level of aspecific activation are also responsible for the presence of incorrect activity on compatible trials--that is, errors occur

provide a description of processes that occur during the foreperiod at a time when no overt behavior is available. Second, we can obtain a precise description of the stimulus evaluation process by looking at speed-accuracy trade-off functions for trials with different P300 latencies. This will enable us to describe in detail the differences in evaluation between compatible and incompatible displays. Together, these two psychophysiological approaches should provide us with a detailed description of two processes that are determinants of the final overt response - stimulus evaluation and aspecific activation and their inter-relationship.

REFERENCES

Ashby, F.G., & Townsend, J.T. Decomposing reaction time distributions. Pure insertion and selective influence revisited. J. Math. Psychol., 1980, 21: 93-123.

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W., & Donchin, E. An ERP/EMG/RT approach to the continuous flow model of cognitive processes (in preparation).

Donchin, E. Surprise! ... Surprise? Psychophysiology, 1981, 18: 493-515.

Donchin, E. The relevance of dissociations and the irrelevance of Dissociationism: a reply to Scwartz and Pritchard. Psychophysiology, 1982, 19: 457-463.

Donchin, E., Coles, M.G.H., & Gratton, G. Cognitive psychophysiology and preparatory processes: A case study. In S. Kornblum and J. Requin (Eds.), Preparatory States and Processes. Erlbaum. Hillsdale, 1984: 155-178

Donders, F.C. On the speed of mental processes. In W.G. Koster (Ed. and trans.), Attention and Performance II. North-Holland, Amsterdam, 1868(1969): 412-431.

investigators. The polygraph is interfaced directly with a computer, thus making hand-scoring of polygraph records unnecessary.

The connection between subject and polygraph is achieved via wires or cables (leads). Their function is merely to transmit electrical activity to and from the subject (electrodes) or to and from the transducers. Each psychophysiological measure is processed by a separate channel of the polygraph. Each channel contains a device which is directly connected to the subject or transducer (sometimes called a "coupler") and an amplifying system. The amplifying system is generally the same for all channels. Most manufacturers of polygraphs supply a variety of couplers each of which is specific for the measurement of a particular psychophysiological function. Below we review some general characteristics of these couplers/amplifiers.

2.2.1 Amplifiers

The most elementary function of the polygraph is to magnify psychophysiological signals. Amplifiers fulfill this function by increasing the magnitude of the input voltage by a factor of up to 500,000. Following amplification, the signal should have an amplitude on the order of about +/- 1 V to be compatible with either the graphical read out system of a polygraph or the analog-to-digital converter of a computer (see below).

The size of the amplification factor will depend on the size of the input signal. For example, the magnitude of the EKG signal is about 1 mV, while that of the EEG is about 50 microvolts. Thus, the amplification factor for these two measures might be 1000 and 20,000 times respectively.

To ensure that the amplifier is performing the appropriate magnification it is important to pass calibration voltages of known amplitude through the amplification system.

amount of light backscattered (if source and receiver are on the same side). Depending on the characteristics of the receiver, variations in the amount of transmitted or backscattered light are converted into variations in electrical current or electrical resistance. In the latter case, a bridge circuit must be used to convert resistance change to voltage change.

In this section we have considered devices that are used to convert the activity of physiological functions into electrical activity. Note that the transducer can only operate on that aspect of the function it was designed to detect. The function may have many manifestations, only one of which is detected by the transducer. Furthermore, the transducer will not differentiate between activity that is caused by the function of interest and that caused by extraneous events. For example, respiration strain gauges will be sensitive to all forms of movement - not just those attributable to respiration. Thus, however well a transducer is designed and positioned, it will be blindly faithful in converting what it "sees" into electrical activity. With these caveats in mind, we can now turn to the system that scales these diverse voltage x time functions to a common format.

2.2 The Polygraph

"Polygraph" is a generic name for a device which amplifies, shapes, and records psychophysiological functions. Although polygraphs come in different shapes and sizes, they have a number of common features: amplifiers, bridge circuits, integrators, rate devices, analog filters, and a graphic read-out facility. The increasing use of computers in psychophysiological research has made the last item redundant for many

mechanical activity. The respiration belt and the strain gauge plethysmograph both rely on the fact that mechanical changes occur with variations in the activity of the function. Respiration may also be measured using a less direct mechanical procedure (the respiratory spirometer) which converts the changes in air flow which occur during respiration into mechanical changes. In other cases, the fact that the function is manifested in changes in the optical quality of tissue is used (e.g. the photoplethysmograph).

The task of the transducer is to convert the mechanical or optical manifestation of the function into an electrical function. With primary and secondary mechanical systems, the conversion can be made to electrical resistance using a strain-gauge. The prototypical strain gauge is a plastic tube filed with mercury. Variations in the length and cross-section of the tube, resulting from stretching, are associated with changes in resistance of the tube. Appropriate placement of the strain gauge ensures that variations in the resistance of the strain gauge are due to variations in the function of interest. Using a suitable bridge circuit (see below), these changes in resistance are then converted into changes in voltage.

Other functions which can be monitored using the resistance principle include temperature. In this case, a thermistor is used whose resistance changes with temperature.

With optical systems, the need is to convert variations in the optical properties of tissue which are associated with vascular events into electrical activity (see Jennings, Tahmoush, & Redmond, 1980). In all optical systems, there are two elements, a light source and a receiver. Activity at the receiver depends either on the amount of light transmitted (if source and receiver are on opposite sides of the tissue) or on the

most recording applications. The primary issue is that there should be chemical overlap among electrolytes at each interface of material. Silver chloride on the electrode surface plus sodium chloride in the jelly creates an appropriate sequence of electrolytes between metal electrode and skin. As noted above, measurement of EDA presents a special set of problems, since the behavior of the system itself can be influenced by the electrolyte. Venables and Christie (1980) present a detailed discussion of the problems of electrolyte with special reference to the measurement of EDA.

The particular characteristics of electrodes, skin preparation, and electrolyte are chosen for one reason--that is, to provide faithful transmission of the electrical activity manifested at the skin to an amplifying system (in a polygraph) where the electrical activity can be magnified. The selection of these characteristics is based on the requirement that whatever reaches the amplifying system should consist of no less and no more than what actually exists at the skin. Note that the activity at the skin may not always represent the activity of interest. Electrodes cannot discriminate among brain electrical activity, muscle electrical activity, or the electrical activity associated with eye-movements. For this reason, care must be exercised in ascribing a cause to the electrical activity recorded using electrodes. We will consider how this activity is treated by the polygraph after we have discussed the second type of subject attachment.

2.1.2 Transducers

Many physiological functions of interest are not directly manifested in electrical activity at the skin surface. The activity may appear in a number of different ways. First, it may appear directly as

the electrodes are used to apply small constant voltages or currents to the skin in order to quantify properties other than surface voltage.

Electrodes customarily are small metallic discs or disc shapes which are attached to the surface of the subject's skin. Placement will depend on the function of interest. Attachment to the subject is generally accomplished through the use of double adhesive collars which stick to both the electrode and the subject. However, if the electrodes are to be placed on an area that is hairy (e.g., the scalp), then either a glue (e.g., collodion) or a rubber cap may be needed to hold the electrodes in place.

The most critical aspect of the electrode is that it is electrically stable. It should be both inert (have no inherent electrical activity) and non-polarizable (be unaffected by continued exposure to current flow). For all functions, the electrode material of choice is currently silver/silver chloride (silver chloride surface surrounding a solid silver base).

Prior to electrode attachment, the skin is generally cleaned with a mild solvent such as acetone. With EDA, however, the measure itself can be influenced by the method of cleaning. Venables and Martin (1967a) report that, while acetone, ether, and distilled water do not effect EDA, soap and water lower conductance and raise resistance. To eliminate the possibility of between subjects variations due to the method of cleaning, these authors advise standardizing procedures across subjects.

Contact between electrodes and skin is maintained by a jelly or paste. For all functions, it is desirable that the jelly be chemically compatible with the skin. For this reason, electrolytes containing NaCL or KCl are generally used, preferably in concentrations that correspond to those found on the skin. Although commercially available electrolytes do not always satisfy this last requirement, they are usually judged to be acceptable for

2. Deriving Voltage x Time Measurement Functions

In this section, we consider the sequence of events (and associated equipment) which transpires between variations in the activity of a physiological system in a human subject and the derivation of the voltage x time functions which represent this activity. This will be a brief review. More detailed treatments can be found in other chapters in this volume, and in Brown (1967), Martin and Venables (1980), Stern, Ray, and Davis (1980), and Venables and Martin (1967b).

2.1 Attachments to the Subject

We may distinguish here between two classes of attachments. First, there are those that are used when the investigator is interested in the activity of a physiological function which manifests itself in variation in electrical activity that can be measured on the surface of the skin. Secondly, there are those that are used when the activity of the function of interest is manifested in a non-electrical fashion. We will consider these two separately.

2.1.1 Electrodes

Electrodes are used when the activity of the psychophysiological function of interest can be detected in the form of electrical activity at the surface of the skin. Measures of the electroencephalogram (EEG), the electromyogram (EMG), the electro-oculogram (EOG), the electrocardiogram (EKG), and electrodermal activity (EDA) all require the use of electrodes. In most cases the electrodes merely constitute an interface between the subject and amplification equipment (see below), although for some measures of EDA (skin conductance and resistance)

This creates some special problems when appropriate values for parameters of cardiovascular functioning in real, rather than cardiac, time are derived (e.g., Graham, 1978).

In spite of occasional esoteric factors that tie particular analytical techniques to particular measures, we propose that, in general, such ties are based on little more than historical accident. Adherence to a technique for the sake of history may be constricting, and part of the aim of this chapter is to encourage a break with tradition. We hope that investigators will consider enriching their analytic repertoires by including techniques that are either customarily employed in other branches of psychophysiology or not currently in use. In this way, the range of questions that can be answered with respect to a given psychophysiological function can be extended.

Our emphasis on the potential generality of analytical techniques should not be taken to mean that we think that specific measurement techniques are unimportant. Other chapters in this volume discuss the measurement techniques that are typically used in the recording of different psychophysiological functions. Furthermore, for the sake of completeness, we briefly review different approaches to psychophysiological measurement in Section 2 (below). However, the bulk of this chapter will be devoted to a review of analytic techniques. We present, in detail, two classes of analytic techniques: time domain and frequency domain. Selection between these two classes, and among the different techniques within each class, is dictated by the questions asked by the investigator. Thus, we will not only describe the different techniques but also point to those questions which the techniques are best suited to answer.

PRINCIPLES OF SIGNAL ACQUISITION AND ANALYSIS

1. Introduction

This chapter describes various techniques that can be used to analyze psychophysiological measures. We approach this description with the assumption that there is a general set of principles that can be applied to any psychophysiological measure, regardless of its origin. We justify this assumption by the observation that all psychophysiological signals are reducible through appropriate measurement techniques to voltage x time functions. Note that we are distinguishing between measurement procedures and analytic procedures. The former may be peculiar to a specific function. For example, the measurement of electroencephalographic activity requires the use of two electrodes and amplifiers to derive voltage x time functions, where the voltage represents some simple transformation of the voltage difference between the two electrodes. Measurement of electrodermal activity (skin conductance), on the other hand, requires not only the use of two electrodes and an amplifier, but also some kind of bridge circuit to translate the variations in skin conductance beneath the electrodes into a voltage x time function.

Although measurement procedures may be "special" in the sense that each psychophysiological function has its own procedure, analytic procedures need not be special because they are all applied, in the end, to a voltage x time function. Of course, traditionally some functions have been associated with specific types of analyses. In some cases, this tradition is justified because of the special characteristics of the psychophysiological function in question. For example, for most measurements of the cardiovascular system, data are available only at each heart beat, and not continuously.

9. Footnotes 125

10. Figure Legends 126

4.2 Time Series Analysis: Definitions and Methods 72
4.2.1 The Definition of a Time Series 72
4.2.2 Time-Domain and Frequency Domain Methods: An Overview 72
4.2.3 Other Frequency-Domain Methods 75
4.2.4 Time Series Statistics: Methods to Partition Variance 77
4.3 Constraints and Limitations of Sampling Procedures 80
4.3.1 Physiological Activity: Continuous Processes 80
4.3.2 Physiological Activity: Discrete Processes 81
4.3.3 Physiological Activity: Point Processes 84
4.4 Conclusion 86
5. Inference Testing 87
5.1 Introduction 87
5.2 Univariate Analysis of Variance (ANOVA) 88
5.2.1 Analysis of Covariance (ANCOVA) 89
5.2.2 Change Scores and the Law of Initial Values (LIV) 92
5.2.3 The Assumption of Homogeneity of Covariance 93
5.2.4 Power of the F-test in ANOVA 96
5.3 Multiple Regression/Correlation (MRC) 98
5.4 Multivariate Techniques 101
5.4.1 Multivariate Analysis of Variance (MANOVA) 101
5.4.2 Canonical Correlation Analysis 102
5.5 Nonparametric Tests 103
5.6 A Few More Caveats 105
6. Conclusion 108
7. Reference Notes 109
8. References 110

3.2.3.4 Discriminant Analysis 27
3.2.3.4.1 Introduction 27
3.2.3.4.2 Linear Stepwise Discriminant Analysis (LSDA) 29
3.2.3.4.3 Applications of LSDA 32
3.2.3.4.4 Evaluation of Discriminant Analysis 34
3.2.4 Digital Filtering 36
3.3 Data Reduction Techniques 38
3.3.1 Introduction 38
3.3.2 Peak Measurement 40
3.3.3 Area Measurement 44
3.3.4 Principal Components Analysis (PCA) 46
3.3.4.1 Introduction 46
3.3.4.2 Appropriate Experimental Design 50
3.3.4.3 Selection and Computation of the Input Matrix 51
3.3.4.4 Extraction of Principal Components 53
3.3.4.5 Rotation of Component Loadings 55
3.3.4.6 Inference Testing 57
3.3.5 Summary Comparison of Data Reduction Techniques 57
3.4 Spatial Analysis 59
3.4.1 Introduction 59
3.4.2 Isopotential maps 60
3.4.3 Univariate and Multivariate Approaches to Spatial Analysis 62
3.4.4 Multivariate Approach to Spatial Analysis 63
3.4.4.1 An Application: The Vector Filter 65
4. Data Analysis in the Frequency Domain (by Stephen W. Porges) 69
4.1 The Description and Partitioning of Variance 69

Table of Contents

Page

1. Introduction 1

2. Deriving Voltage x Time Measurement Functions 3

2.1 Attachments to the Subject 3

2.1.1 Electrodes 3

2.1.2 Transducers 5

2.2 The Polygraph 7

2.2.1 Amplifiers 8

2.2.2 Bridge Circuits 9

2.2.3 Analog Filtering 9

2.2.4 Analog Integration 11

2.2.5 Rate Devices 12

2.3 Computer Access to Voltage x Time Functions 13

2.3.1 Digital Input and Analog-to-Digital Conversion 13

2.3.2 Distributed Processing: Remote Data Acquisition 14

3. Data Analysis in the Time Domain 16

3.1 Introduction 16

3.2 Signal Extraction Techniques 17

3.2.1 Signal Averaging 18

3.2.2 Removing Systematic Noise 19

3.2.3 Pattern Recognition 22

3.2.3.1 Introduction 22

3.2.3.2 Cross-Correlation 24

3.2.3.3 Woody Adaptive Filter 26

PRINCIPLES OF SIGNAL ACQUISITION AND ANALYSIS*

Michael G. H. Coles, Gabriele Gratton,

Arthur F. Kramer, and Gregory A. Miller

Cognitive Psychophysiology Laboratory
Department of Psychology
University of Illinois
Champaign, Illinois 61820

With a Section on

"Data Analysis in the Frequency Domain"**
Stephen W. Porges, Department of Psychology,
University of Illinois
Champaign, Illinois

Note: The order of authors is alphabetical. No ranking is implied.

*Acknowledgements: The preparation of this chapter was supported in part by the Air Force Office of Scientific Research (Contract No. F49620-79-C-0233); the Environmental Protection Agency (Contract No. EPA CR 808974-02); School of Aerospace Medicine, Brooks Air Force Base (Contract No. F33615-82-C-0609); and the Defense Advanced Research Projects Agency (Contract No. MDA903-83-C-0017).

**Acknowledgements: The preparation of this section was supported, in part, by Research Scientist Development Award K02-MH-0054 from the National Institute of Mental Health and by grant HD 15968 from the National Institutes of Health.

To appear in: Coles, M. G. H., Donchin, E., and Porges, S. W., Psychophysiology: Systems, Processes, and Applications. Vol I: Systems. New York: Guilford Press, in press.

Pachella, R.G. The interpretation of reaction time in information processing research. In B.H. Kantowitz (Ed.), Human Information Processing: Tutorials in Performance and Cognition. Erlbaum, Hillsdale, 1974: 41-82.

Pieters, J.P.M Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. Psychol. Bull., 1983, 93: 411-426.

Posner, M.I. Chronometric Explorations of Mind. Erlbaum, Hillsdale, 1978

Rohrbaugh, J.W., Syndulko, K., & Lindsley, D.B. Brain components of the Contingent Negative Variation in humans. Science, 1976, 191: 1055-1057.

Sternberg, S. The discovery of processing stages: Extensions of Donders method. In W.G. Koster (Ed.), Attention and Performance II. North-Holland. Amsterdam, 1969a: 276-315.

Sternberg, S. Memory scanning: Mental processes revealed by reaction time experiments. Am. Scient., 1969b, 57: 421-457.

2.2.2 Bridge Circuits

As we have seen, most transducers represent psychophysiological activity in the form of resistance changes. For this reason, a critical function of the polygraph is to measure resistance change and to convert it to voltage change. This is accomplished through the use of a bridge circuit, which can be as simple as a few resistors arranged in a special way (Malmstadt, Enke, & Crouch, 1974). A bridge circuit provides constant current to the transducer. As the resistance of the transducer changes, so the voltage across the transducer changes. This voltage change is then amplified (see above).

Bridge circuits are also used in the measurement of two complementary forms of electrodermal activity, skin conductance and skin resistance. In this case, either a constant voltage or constant current is imposed on the subject, and the bridge measures variations in current or voltage which correspond, respectively, to variations in conductance or resistance. Because this procedure involves the imposition of external electrical activity on the subject, safety is a critical factor. However, the procedure is now reasonably standardized (see Fowles chapter).

2.2.3 Analog Filtering

As we have mentioned, the task of the electrodes and transducers is to convey to the polygraph a faithful representation of the electrical or other activity associated with a psychophysiological function. In some cases, the signal so conveyed may be filtered by the polygraph, either because it contains artifact or because it contains aspects of the psychophysiological signal which are of no interest to the investigator.

For the purposes of describing the principles of signal modification or "signal conditioning", the signal is considered as being comprised of different frequencies. Thus, some of these frequencies may be artifactual (due to sources outside the subject or to activity of other, irrelevant functions), while others may simply be of no interest.

For example, a common source of artifact in psychophysiological measurement is 60 Hz (or 50 Hz) activity from standard electrical equipment. This artifact can be minimized by the use of a "notch" filter set at 60 or 50 Hz, which attenuates activity at this frequency while permitting activity at higher or lower frequencies to pass.

Other filters attenuate activity above or below specified frequencies (low-pass and high-pass filters). For example, in EEG recording the investigator is generally interested only in activity below 40 Hz. Thus, a low pass filter set at 40 Hz can be used. The EKG consists of frequency components between .05 Hz and 80 Hz (Strong, 1970). If the investigator merely wants to detect the R-wave (e.g., to measure interbeat interval), a high-pass filter set at 10 Hz can be used. The high pass filter attenuates slow shifts in the EKG signal that may be due to electrodermal activity or some other unwanted activity.

The various types of electronic circuitry which typically serve as filters can be characterized by their "time constant". The value of the time constant of a high-pass filter is the time for a given sustained input to the circuit to be attenuated to 63% of its original value. While all analog filtering circuits have a time constant characteristic, in practice the concept is associated primarily with that portion of a circuit which serves as the high-pass filter. Some correspondences between time constant

and filter cutoff frequency (-3dB) are as follows ($F = 1 / (2*\pi*TC)$, TC = time constant):

Time Constant (sec)	Frequency (Hz)
10	.016
5	.032
1	.159
.3	.531
.01	15.915

Of course, it is imperative that great care be taken in the use of filters. The investigator does not want to distort the signal of interest. Filters are useful when the characteristics of the unwanted aspects of the signal do not overlap the wanted aspects. The problems that occur when there is overlap, and the solutions to these problems, will be discussed below (Section 3).

The "analog" filters briefly discussed here are electronic components placed in-line during initial recording of continuous signals, often within the amplifier chassis. Their chief advantages are simplicity and speed. Their disadvantages are that they introduce a phase shift into the signal and that in a particular polygraph they are typically limited to a few settings. Analog filters must be distinguished from digital filters, which are algebraic manipulations of discrete (digitized) signals after recording is complete. Digital filters, discussed in Section 3.2.4, can be constructed without phase shift and with any filter characteristics.

2.2.4 Analog Integration

For some physiological signals, particularly EMG, the investigator is not so much interested in the frequency characteristics of the signal as in the overall amplitude-frequency activity in the signal. Analog integrators provide this measure by first rectifying the signal and then converting the area under the rectified record into a smoothed analog

voltage (rectification involves removing or inverting either the positive or negative portion of an AC signal). The resulting voltage x time function will depend on both the amplitude and frequency of the input signal at any point in time. Because analog integration is normally accomplished with an in-line electronic circuit which is essentially a low-pass filter (smoothing out rapid peaks but preserving average amplitude), different integrators are appropriate for different physiological signals, depending on the frequency characteristics of the signal and the time constant of the integration circuit. Furthermore, the output of such an analog circuit lags the input, again introducing the issue of phase shift. When the frequencies of interest are high relative to the time resolution needed, as in EMG recording, this lag is inconsequential.

2.2.5 Rate Devices

With some physiological functions, the measure of interest is the rate at which some event occurs, rather than the level of activity. For example, with heart rate (HR), the investigator is concerned with the rate at which 'R' waves are observed in the EKG, rather than with voltage characteristics of the EKG waveform itself.

To accomplish this measurement, most polygraph manufacturers offer rate devices (cardiotachometers) which convert inter-event intervals into an analog signal whose amplitude varies with rate. In some implementations, the conversion is made through a circuit which first detects an 'R' wave, then allows a capacitor to be charged until the next 'R' wave is detected, at which time the capacitor is discharged. The voltage discharged by the capacitor will vary as a function of the duration of the charging period, and hence will be proportional to the inter-beat interval (and inversely

information (0 or 1), the output has a large number of possible values. For example, a 12-bit A/D converter can output 4096 different values, depending on the voltage input at the time of sampling. Such resolution is essential for measurement of signal amplitude. The sampling intervals used vary as a function of the particular measure. For example, for the auditory brain stem response the intervals are typically 20 microsec (sampling rate of 50 kHz), while for respiration the intervals may be as long as 1 sec (1 Hz). Choice of sampling interval (or sampling rate) is dictated by the expected period or frequency characteristics of the measure in question. The slowest acceptable sampling rate is twice the highest frequency present in the data. A slower sampling rate will provide a distorted digital representation of the analog input (this issue is elaborated further in Section 4). A good rule of thumb, then, is to err on the conservative side and sample at least 2-5 times the expected frequency.

The output of the A/D converter, now a discrete voltage x time function, is fed directly to the computer. While logically distinct from the computer itself, circuitry such as Schmitt triggers, digital input interfaces, and A/D converters are typically integrated electronically into the computer enclosure.

2.3.2 Distributed Processing: Remote Data Acquisition

Given the low price and small size of current microprocessors, laboratory equipment manufacturers have begun to offer "smart" laboratory products which perform the continuous-to-discrete conversion external to the computer and its associated A/D converters, etc. Data are then passed to the computer in highly palatable form--as the same 8-bit characters that video display terminals send. Thus, the traditional configuration of "dumb"

proportional to the rate). Note that the level of the output of the rate device (a voltage x time function) will depend on the previously completed inter-beat interval. Thus, the output will lag the input.

2.3 Computer Access to Voltage x Time Functions

2.3.1 Digital Input and Analog-to-Digital Conversion

With the development of computers, the possibility of automatic scoring of physiological data has become a reality. But, before a digital computer can apply the appropriate scoring algorithms, the data must be presented in a palatable form--a set of digitized (i.e., discrete) values. However, the voltage x time functions we have decribed are inherently analog (i.e., continuous) functions. The requirement, then, is to convert these analog functions into digital representations. Some types of physiological activity are easily represented digitally. For example, while the EKG is a continuous voltage, the occurrence of its R-wave component is easily approximated digitally as a "one" in a series of "zeros". Simple electronic circuitry between polygraph and computer, such as a Schmitt trigger, readily converts the analog EKG input signal to such a digital output signal. Thus, a continuous voltage x time signal is converted to a discrete voltage x time signal. This method is more accurate than, and obviates the need for, a cardiotachometer rate device, describe above (Section 2.2.5).

More elaborate conversion circuitry is required when more information about the continuous input function, than the mere occurrence of an event, must be represented in the discrete output function The term "analog-to-digital" (A/D) converter is normally reserved for such circuitry, which produces a series of numerical values which are discrete samples of voltage level from a continuous input. Rather than merely one bit of

equipment plus a dedicated laboratory computer (with central A/D converter, etc.) can be replaced with "smart" equipment plus a simpler, general, multipurpose computer.

The investigator should, of course, consider the growing variety of configuration options in laboratory equipment when developing a new measurement capability. The point is that across these diverse options all psychophysiological data, whether written on polygraph paper or handled by the most elaborate microprocessor network, can be treated as a voltage x time function, a series of voltage levels in time--a voltage "time series".

3. Data Analysis in the Time Domain

3.1 Introduction

This chapter will distinguish analytic techniques applied to data in the time domain from those applied in the frequency domain (see Section 4). Psychophysiologists intend to monitor the activity of some internal structure manifested as a signal conveyed to the body surface by some functional channel. This signal is combined with "noise" coming from other internal and external sources. In many cases the extraction of the signal from the background noise is a very challenging task.

In the case of data in the time domain, the signal is typically a phasic, non-repetitive feature of the time series recorded at the surface which is assumed to reflect the activity of a specific internal structure. Important characteristics of this feature commonly are its restriction to a particular time epoch in the record, and its variability in latency. Since the signal of interest contributes only part of the variability observed in the time domain, we refer at it as a "component". This component constitutes the target of the signal extraction procedure.

Since signal components are in most cases embedded in noise, the first task for the data analyst is to extract the signal from its background. To accomplish this task, the signal must be defined.

Signal extraction techniques differ in the way in which they define components. The choice of an extraction technique implies a model of the signal, including a specification of its distinctive features and how these interact to produce the waveforms (time series) which are actually recorded. For instance, a model of event-related potentials (ERPs) could define a component as a deflection of the EEG trace time-locked to a stimulus, with a specific latency and scalp distribution that "summates" with other

components and with noise to produce the waveforms recorded at the scalp. Alternatively, EDA components are deflections of the skin conductance trace with some shape and latency following an eliciting stimulus. A cardiac cycle can be identified by means of a distinctive feature (R-wave), or by its general waveshape, referred to the spatial location used for the recording. Analogous definitions can be given for any component of interest for the psychophysiologist. Specific component models are often highly controversial. Nevertheless, the procedure adopted to extract the signal from the noise in which it is embedded necessarily depends on some kind of model. Therefore, in the present discussion we will pay particular attention to models of signal and noise implicit in different signal extraction techniques.

Once the signal component is defined, the amplitude, latency, or spatial distribution of the raw data can be quantified. These quantification techniques depend on the definition of "components" used for extracting the signal. In many cases, these two stages of data analysis (quantification and signal extraction) constitute a single process. However, the logical distinction between quantification and signal extraction should be kept in mind throughout this chapter.

3.2 Signal Extraction Techniques

The remainder of Section 3 provides a brief sample of the many ways to process the basic voltage x time function. This review is divided into techniques for signal extraction, for data reduction, and for spatial analysis. In fact, since a given technique may serve several such functions, such a division is necessarily somewhat arbitrary.

3.2.1 Signal Averaging

Since the psychophysiological signal is often obscured by noise, many techniques have been proposed to amplify selectively the information of interest for the psychophysiologist. A number of techniques assume that the signal can be differentiated from the noise on the assumption that only the signal is temporally related to an external marker event. Such procedures therefore define the signal as everything in the recording which is time-locked to an external event. All other variability, not time-locked to the external event, is considered noise. This definition is particularly useful when studying perceptual and motor processes. In this case, the relevant external events are readily identifiable, and the temporal relationship of the external event and the internal process is assumed to be constant. The basic procedure consists of the repetition of a large number of essentially identical trials. Through superimposition, or averaging, of the single trials the constant psychophysiological response (signal) to the stimulus remains constant, while variability not consistently related to the external event averages to zero.

The superimposition technique consists simply of overlapping on a plotter the trace for each of the single trials. It can be also obtained with a storage oscilloscope, by triggering the display sweep at each presentation of the stimulus. Since superimposition does not require high-speed computing facilities or analog-to-digital conversion, it was extensively employed in 1950s. An advantage of superimposition is that it portrays the range of variability of the single trials. However, it is fairly difficult to detect small potentials, or small differences in amplitude between conditions, by means of this technique. Thus,

superimposition is more appropriate when measuring latency than amplitude. However, in recent years it has been replaced by averaging techniques.

In averaging, the values obtained at each time point are averaged across trials. To employ this algebraic technique, it is, of course, necessary to transform the signal obtained from the amplifier from analog to digital format.

The advantage of averaging over superimposition is the "cleaner" waveform which averaging produces. This expresses the "central tendency" of the sample of trials examined and corresponds to the best statistical estimate of the signal. It is easy to compute the point-by-point standard deviation or range in parallel with the averages, in order to have more complete information about the data.

In principle, averaging can extract an arbitrarily small signal relative to background noise amplitude, if a large number of invariant trials are averaged. The noise will be reduced as a function of the square root of the number of trials. For example, the brainstem auditory ERP, typically less than 1.0 microvolt, may require several thousand trials.

However, averaging is vulnerable to violations of its assumptions of specifiable external stimulus and invariant response latency and morphology. Particularly when the investigator suspects cross-trial inconsistency in the signal, averages must be interpreted cautiously.

3.2.2 Removing Systematic Noise

Most signal extraction techniques have been developed in order to deal with the problem of <u>random</u> noise. Consequently, they are often insufficient in the case of <u>systematic</u> noise. In fact, these techniques generally assume that "noise" is that part of the variance that is not

systematically related to the experimental variables. Of course, this corresponds to the definition of random noise. However, some of the noise present in the data can be systematically related to the experimental variables. We label this "systematic noise".

In the presence of systematic noise, two important points must be kept in mind. First, the signal must be defined in a more restricted way than simply as "everything related to experimental variables". An example is given by Event-Related Brain Potentials (ERPs), where, for a component to be considered a signal, it is not sufficient that it is systematically related to the eliciting event. It is also necessary that it be generated by the brain. Therefore, a systematic ocular potential, recorded at the scalp, does not constitute an ERP component, but systematic noise. This kind of systematic noise is commonly called "artifact".

A second important point concerns the difficulty of dealing with systematic noise by means of traditional signal extraction techniques. A procedure usually adopted to reduce artifact in recording is filtering. There are many ways of filtering data, the most common being frequency filtering. This kind of filter is discussed elsewhere in this chapter (see Sections 2.2.3 and 3.2.4). However, frequency filters are sometimes insufficient for handling artifacts in the data. This is especially the case when signal and artifact have similar frequencies. Eye movement artifact in brain ERPs is an example of this problem.

Fortunately, artifacts are sometimes recognizable by their specific features. These features may be evident in the data themselves or in a recording from electrodes placed near the source of the artifact. In either case, the artifact can be detected (by visual inspection or by some automatic procedure) and the associated record discarded from subsequent

analysis. However, although this is a common procedure, such loss of data is not always affordable (Gratton, Coles, & Donchin, 1983).

For this reason, procedures have been developed in order to compensate for artifact. They are based on the possibility of inferring the effect of the artifact on the records at a certain spatial location from data obtained from a location close to the source of artifact. Data of the latter type may be considered "pure" measures of the activity of the "artifact generator". The remainder of this section will describe a recently developed procedure of this type.

This procedure, proposed by Gratton, et al. (1983), represents an example of an artifact compensation technique. It assumes that the effect of an eye movement on the potential recorded at any scalp location (EEG) can be inferred from activity recorded at a location close to the eyeball (EOG). In order to make this inference it is sufficient to know how much a signal recorded at the ocular electrode "propagates" to the scalp location under study. Previous researchers (e.g., Corby & Kopell, 1972; Overton & Shagass, 1969; Weerts & Lang, 1973), have demonstrated that not all ocular potentials propagate to the scalp in the same way. In particular, potentials generated by eyeblinks propagate less than potentials generated by saccadic eye movements.

Accordingly, the proposed eye movement correction procedure (EMCP) distinguishes between time points in the record during which eyeblinks occur (detected by means of a pattern recognition technique; see Section 3.2.3) and time points in which saccadic eye movements occur. Separate propagation factors are then computed for blinks and saccades.

The propagation factors are computed by means of a least squares regression technique. However, as noted above, ocular artifacts can be consistently related to some external events. Since brain potentials (ERPs) can also be elicited consistently by external events, spurious relationships can affect the computation of the correction factors. Therefore, the averaged EOG and EEG traces are subtracted from the single trial records before the correction factors are computed. In this way the propagation factors are computed on that portion of the variance of the EOG and EEG recordings that is not related to the external event. The propagation factors are then applied to the original data to correct for the ocular artifact. A schematic representation of the procedure is presented in Figure 1.

Although some inaccuracy is present (involving mainly the invariance in

----

Insert Figure 1 About Here

----

time of the EEG and EOG response to the external event, and the difference between the propagation factor for upward and downward eye movements), tests presented by Gratton et al (1983) indicate that this procedure effectively compensates for the ocular artifact.

3.2.3 Pattern Recognition

3.2.3.1 Introduction

Signal averaging techniques (see Section 3.2.1) are particularly useful in separating small signals which are time-locked to an external event from background noise which is not time-locked to the external event. However, in many cases, the assumption of invariance of

3.2.4 Digital Filtering

Digital filters have been little used explicitly in psychophysiology. They constitute an interesting contrast with analog filters (Section 2.2.3). A digital filter is most easily described by example. Conceptually, perhaps the simplest digital filter consists of replacing each value in a time series with the average of that number, the number preceding it, and the number following it. Such a common smoothing operation is a rudimentary low-pass filter, in that high-frequency components are reduced.

Specific digital filters vary along several dimensions, which determine the bandpass characteristics and computational speed of each filter. In the above example, three weights are used, each having a value of 1/3. Somewhat less smoothing is accomplished if a different set of weights is used: 1/4, 1/2, 1/4. Alternatively, smoothing is also altered if the number of weights (the "window width") is changed to 5, each weight perhaps being 1/5. If the number and values of the weights are held constant but the time interval between data points is changed, the filter will again have different characteristics. A final choice is whether to apply the weights recursively--i.e., after applying the filter at point T, does the filter applied to point T+1 employ the unfiltered T (non-recursive) or the filtered T (recursive) in computing the filtered T+1?

Clearly, psychophysiologists routinely manipulate their data algebraically in ways which constitute digital filtering. Even the computation of a mean of N values can be seen as (1) assigning each value a weight of 1/N, (2) applying the filter to the mid-point value in the time series by summing the weighted values, and (3) discarding all but the "filtered" mid-point value in the time series. What is typically not

amplitude and latency of the voltage x time components vary across trials. The weighting coefficients derived in the process of discriminant analysis provide information which can be interpreted in terms of the voltage x time components. Thus, components derived in the PCA procedure can be compared with the time points selected in the discriminant analysis procedure to give the investigator an indication of the important features in the data set. Although the initial calculation of the discriminant function is computationally costly, its application is relatively simple. In most cases it requires only the multiplication and summation of a few variables x weighting coefficients.

There are also several disadvantages to discriminant analysis. The need for an independent basis for grouping voltage x time functions can be problematic in some cases, particularly during exploratory data analysis, in which hypotheses are weak or nonspecific. A second problem is that a useful discriminant function can be calculated only if the groups differ significantly. Finally, the need for cross-validation of the discriminant function imposes additional requirements on the investigator. It is preferable that sufficient data be collected so that the discriminant function can be computed with one set of data and validated on another.

As with other analytic techniques discussed in the present chapter, discriminant analysis cannot be profitably employed without consideration of its limitations and assumptions. However, correct application of the discriminant analysis procedure can produce valuable information for the psychophysiologist.

waveforms in a two-tone discrimination task (Squires et al, 1976). The

----

Insert Figure 3 About Here

----

investigators calculated discriminant scores for fifth-order sequential stimulus patterns to demonstrate the effect of sequence on the amplitude of several components in the ERP waveform. As can be seen from the figure, the discriminant scores obtained in the experiment closely paralleled the sequential structure of the task. Thus, the discriminant tree diagram provides another means of analyzing the fine structure of subjects' behavior.

Discriminant analysis may also be employed to evaluate the degree of resemblance of a single trial to the average of one group or another. In the case of voltage x time functions collected in a psychophysiological experiment, the investigator may wish to know how well a single function resembles the average of one of several groups. This information is provided by the discriminant score.

3.2.3.4.4 Evaluation of Discriminant Analysis

As with other statistical techniques, there are both advantages and disadvantages associated with using discriminant analysis procedures in the evaluation of psychophysiological data. Discriminant analysis provides an objective, quantifiable method of assessing differences in single voltage x time functions both within the training set and across other data sets collected under the same general experimental paradigm. In addition to providing classification information, discriminant analysis also provides an alternative to signal averaging in situations in which the

has been illustrated by several studies which have employed discriminant analysis to assess the group membership of single trial ERPs. In one such study subjects were asked to count covertly the total number of high-pitched tones from a Bernoulli series of high- and low-pitched tones. High tones occurred with a probability of .20 while low tones occurred with a probability of .80. The common finding in this general paradigm is that counted, low-probability events produce larger P300 components than uncounted, high-probability events. Replicating this design, Squires and Donchin (1976) then employed discriminant analysis for the purpose of classifying each single trial ERP as either a high or low probability event. The discriminant function was able to correctly classify 81% of the single trial ERPs. An examination of the averages of correctly and incorrectly classified ERPs (see Figure 2) indicated that the misclassified events resembled the category into which they were classified more closely than

----

Insert Figure 2 About Here

----

they resembled their correct category. These results suggest that some rare stimuli evoked a response characteristic of frequent stimuli and vice versa. That is, rather than erring in its classification of waveforms, discriminant analysis may have identified trials in which the subject erred in classifying stimuli. This example illustrates the heuristic value of the technique in revealing the fine structure of the subject's behavior, which can be obscured by cross-trial signal averaging techniques.

Another example of the use of discriminant analysis in the detailed examination of subjects' behavior is found in tree diagrams of discriminant scores. Figure 3 depicts the discriminant scores obtained from ERP

3.2.3.4.3 Applications of LSDA

The standardized weighting coefficients obtained in the discriminant analysis procedure can provide valuable information concerning the relative importance of the variables employed in the discriminant function. Examination of the weighting coefficients enables the investigator to assess the contribution of each variable in the discriminant function. Large weights, in either a positive or negative direction, denote a substantial contribution of their respective variables to group differentation. In psychophysiological experiments in which the investigator wishes to classify voltage x time functions into two or more groups, the magnitude of the weighting coefficients identifies those features or time points which best differentiate between groups. For example, Horst and Donchin (1980) found that the ERP time points which best differentiated between two pattern-reversal conditions were within the region of the voltage x time functions which were predicted to change as a function of experimental manipulations. Furthermore, these time points were consistent with the components derived in a Principal Components Analysis of the data (see Section 3.3.4 below).

Discriminant analysis also provides a classification rule which best differentiates between the training groups. This classification rule can be applied to other data sets collected in similar paradigms. In this case the investigator is interested in classifying new data according to probability of group membership.

Although the primary purpose of the discriminant function is the correct classification of the greatest possible proportion of cases, useful information may also be obtained from the misclassified cases. This point

alternative techniques. One method, commonly called the "jackknife procedure", removes one case from the training set, computes the discriminant function, and then classifies the case which has been omitted. This procedure is repeated until a discriminant function has been calculated for each of the cases in the data set. Overall classification accuracy is determined by dividing the number of single cases misclassified by the total number of cases contained in the data set. Although the jackknife procedure provides a check on the efficiency of the discriminant function, it does not usually produce results which vary greatly from the original computation. Another cross-validation procedure, the randomization test, is applied to the entire training set. In this instance, however, the cases in the training set are randomly assigned to two groups. A new discriminant function is then computed for these randomly assigned groups. This process is repeated several times and a distribution of discriminant functions is compiled. The distribution provides an indication of the classification results which can be expected with random data, thereby providing the investigator with a basis against which to compare the performance of the original discriminant function.

The linear stepwise discriminant analysis procedure outlined above assumes that the covariance across groups is equal and that noise or error in the data conforms to a normal distribution. In cases in which these assumptions are violated, LSDA will provide less than optimal group discrimination performance. A useful alternative in some of these cases is the quadratic discriminant analysis technique (QDA). The QDA procedure is similar in function to the LSDA technique and has been used successfully in a number of studies (Aunon, McGillem, & O'Donnell, 1982; McGillem et al, 1981; Sencaj, Aunon, & McGillem, 1979).

A separate vector of weighting coefficients will be derived for each of N-1 discriminant functions, N being the number of groups. The discriminant criterion value provides a measure of group differentiation for each discriminant function. The first discriminant function has the largest discriminant criterion value, indicating the dimension of maximal group differentation. The second discriminant function represents the largest group difference not accounted for by the first dimension. Thus, the discriminant criterion value and hence the group differentation accounted for by each discriminant function decreases with successive functions.

Although N-1 discriminant functions can be calculated, they might not all contribute significantly to group differentation. Several procedures are available to test the incremental significance of successive discriminant functions (see Tatsuoka, 1970, 1971). Eliminating discriminant functions which do not contribute significantly to group differentation serves further to reduce the dimensionality of the data set.

As mentioned above, one of the main functions of discriminant analysis is to provide a classification rule which correctly identifies a high proportion of cases. However, the usefulness of the discriminant function is not determined solely on the basis of its classification accuracy with the original data set (training set). Cross-validation is necessary to establish the validity of the discriminant function. When the investigator has a large number of cases available, the most direct procedure is to divide the data set in half, calculate the discriminant function with one half of the data and validate it on the other half. This procedure can also be carried out on a new data set collected under the same general experimental paradigm. If, on the other hand, the investigator has an insufficient quantity of data to perform this procedure there are several

3.2.3.4.2 Linear Stepwise Discriminant Analysis (LSDA)

The most commonly used discriminant analysis procedure for the assessment of psychophysiological data is the linear stepwise discriminant technique (Donchin & Herning, 1975; Horst & Donchin, 1980; McGillem, Aunon, & Childers, 1981; Squires & Donchin, 1976). The goal of the LSDA procedure is the selection of a subset of variables which maximize the between-group separation. The process is analogous to stepwise multiple regression, except that in LSDA the predicted criterion can be a multi-level nominal variable.

The first step is to identify the variable which accounts for the largest proportion of between-group variance. A second variable is then selected which accounts for the maximum proportion of between group variance not already accounted for by the first variable. This successive selection of variables constitutes the stepwise portion of the LSDA procedure. The between-group difference at each step in the procedure is measured by a one-way analysis of variance F statistic, and the variable with the largest F is chosen. Several LSDA computer programs permit the deletion of variables which no longer provide a substantial contribution to group separation as other variables are added (Dixon, 1979; Jennrich, 1977). These variables may later be re-entered if their F value is again adequate. The process of variable selection is terminated when some specified criterion has been met. Criteria commonly employed include: the number of variables already entered, the amount of variance accounted for, or the point at which no further improvement occurs in some criterion (e.g., the U statistic in the BMDP package).

between-group variance while minimizing the within-group variance. In the case of psychophysiological data, when the investigator wishes to discriminate between sets of voltage x time functions, the discriminant function consists of a linear combination of time points x weighting coefficients.

As has been discussed above (Section 3.2.1), signal averaging can serve as a relatively simple method of pattern recognition and signal classification. One might doubt the necessity of employing more complex, multivariate techniques such as discriminant analysis to accomplish the same goal. In many situations, averaging is adequate. In some situations, however, signal averaging will produce misleading results. For example, averaging is inappropriate when substantial, uncontrolled variation in the amplitude of a component occurs. In such cases, discriminant analysis provides a clear advantage over signal averaging procedures since the differential amplitude of the psychophysiological component can become the basis for group classification. In addition to supplementing the signal averaging procedure, discriminant analysis also provides a technique which can be employed in the analysis of single-trial data. This is clearly advantageous when the investigator is interested in the trial-to-trial variation in both psychophysiological and performance measures. For example, the use of discriminant analysis procedures in the evaluation of single trial ERPs has had important theoretical implications. Squires, Wickens, Squires, and Donchin (1976) employed it to construct a quantitative expectancy model of the P300 component of the ERP (see below, Section 3.4.3.2.3).

able to improve the signal-to-noise ratio over a definite limit. Therefore, its reliability under conditions of very low signal-to-noise ratio is questionable.

3.2.3.4 Discriminant Analysis

3.2.3.4.1 Introduction

Discriminant analysis provides a method of discriminating between two or more groups on the basis of systematic differences in the data set. A case classification rule is derived from data whose group membership is known (training set data). This rule is then applied to new data of unknown group membership (test set data). Thus, the discriminant analysis technique requires that the investigator specify *a priori* the groups into which the data are to be classified. Groups may refer to distinct samples of subjects or to distinct classes of events which vary within subjects.

In addition to providing a method of discriminating among groups, discriminant analysis also provides a means by which to reduce the dimensionality of the data. Such a reduction serves to increase the stability of the discriminant composite. Data reduction is accomplished by selecting a subset of the original variables which best discriminates among the groups. These variables are then used in computing a linear combination of weighting coefficients x variables to produce a discriminant score. The pattern of weighting coefficients provides information concerning the contribution of each variable to the differentiation between the groups. The function employed in the computation of the discriminant score is referred to as the discriminant function. The purpose of the function is to provide optimal separation between two or more groups by maximizing the

3.2.3.3 Woody Adaptive Filter

The Woody Adaptive Filter (Woody, 1967) is a particular kind of cross-correlational technique. The term "adaptive" refers to the fact that the template is not established _a priori_, but is extracted by means of an iterative procedure from the data themselves. Each iteration serves to refine the template. This method was originally proposed to identify particular patterns of variation of the EEG recorded in epileptic patients.

The Woody Filter makes use of an adaptive template. Typically, the template used for the initial iteration is the half-cycle of a sine or triangular wave or the average of the unfiltered single trials. Cross-lagged covariances or correlations are computed between each trial and this template. A new template is obtained by aligning the single trials at the lag which gives the maximum cross-correlation. This procedure is then repeated, using the new average as the template, until the maximal values of cross-correlation become stable. Trials where correlations with the template do not reach a criterion (e.g., .3 to .5) at any lag are not used in subsequent template construction and may be discarded entirely from subsequent analysis.

Several studies have been conducted to test the power and reliability of the Woody Filter (Nahvi, Woody, Ungar, and Sharafat, 1975; Woody and Nahvi, 1973; Wastell, 1977). They have concluded that the Woody Filter method is often superior to a simple peak detection technique (see Section 3.3.2). However, the use of multiple iterations has been questioned (Wastell, 1977). In fact, this author reports a decline in validity of the procedure when several iterations are used. Therefore, in contrast with signal averaging, Woody filter (and most auto-correlation techniques) is not

By progressively increasing the size of the lag, a series of correlation values is computed, limited only by the number of the elements of the trial array (a correlation involving too a small number of elements would not be reliable). Then, the maximum value in the series of cross-correlations is selected. The lag corresponding to this maximum value is the one at which the trial maximally "looks like" the template. According to the pattern recognition approach, this is the lag at which the signal is "detected". In most cases, if for a given trial some minimal correlation value can not be reached with any lag, the "signal" is considered to be absent on that trial.

Cross-correlation is vulnerable to two problems. First, the maximum cross-correlation for a particular trial may be unacceptably low. To accept such a trial is to assume that the signal is whatever in the data is least dissimilar to the template. Second, cross-correlation cannot easily handle the presence of multiple components differing in latency. This would constitute a violation of the assumption of invariance of shape of the signal. Cross-correlational techniques should not replace signal averaging techniques in the case of components with fixed latency, particularly when the signal-to-noise ratio is very small.

Notwithstanding these limitations, cross-correlation techniques have the advantage of utilizing the information provided by the <u>whole</u> time-series, thereby increasing the power of the analysis. They are particularly useful in identifying components having variable latencies embedded in large amounts of noise.

sometimes be applied without any previous knowledge of the "pattern" or "feature" to be recognized. Two examples of this kind of technique (cross-correlation and discriminant analysis) will be discussed below in some detail.

3.2.3.2 Cross-Correlation

A fundamental assumption of the cross-correlational approach (Friedman, 1968) is that the waveshape of the signal component to be detected is constant over trials, while the shape of the noise varies randomly from trial to trial. Thus, that portion of the variance which is constant over trials will contribute to the correlation between trials. Because cross-correlational techniques do not assume invariance of the interval between external event and internal process manifested by the signal component of interest, they are applicable even in the absence of any identified external event and potentially more versatile than signal averaging, which assumes signal invariance in both morphology and latency.

Cross-correlational techniques involve the computation of a "cross-correlational series" between a "template" (pre-determined pattern of consecutive points) and any single trial. A cross-correlational series is an array of correlation values between two time series (or within the same series), where one of the time series is progressively shifted by a certain interval (lag). For example, the first correlation index is computed between the elements (a(1),a(2),a(3),...,a(n)) of the template and the elements (b(1),b(2),b(3),...,b(n)) of a given trial. In this case the "lag" between the elements of the template and of the trial is 0. Then a second correlation index is computed between the elements (a(1),a(2),a(3),...,a(n)) of the template, and the elements (b(1+lag),b(2+lag),b(3+lag),...,b(n+lag)).

latency of the signal over trials is untenable, even as a first approximation. In other cases it is impossible to establish an external event to which the psychophysiological signal can be time-locked. Thus, straightforward signal averaging is not always possible.

Pattern recognition techniques can be helpful in these cases. The general assumption underlying such techniques is that the signal is distinguishable from the background noise on the basis of specific features, typically aspects of its waveshape. Two types of pattern recognition techniques may be distinguished: those in which the characterizing features are established *a priori* on the basis of previous data or conceptualizations, and those in which the characterizing features are established *a posteriori* on the basis of characteristics of the data to be processed.

Examples of the first type are the techniques used in psychophysiology to detect the R-wave of the EKG, the phasic electrodermal response, or blinks in the EOG trace. Procedures of this type are specific for a particular psychophysiological measure and are not easily generalizable to other measures. Note also that many of these pattern recognition techniques are used to recognize artifacts (see also Section 3.2.2 on EMCP). Although pattern recognition may be performed simply by visual inspection of the records, for reasons of reliability it is preferable to automate the procedure using hardware devices (e.g., Schmitt triggers) or software algorithms.

Pattern recognition techniques based on standard statistical procedures (e.g., cross-correlational techniques, discriminant and canonical analyses, etc.) usually require the use of high-speed computing devices, since they involve large amounts of computation. However, they have the advantage of being generalizable to many different measurement domains; they can also

discussed when simple digital filters are used are the bandpass characteristics of the filter procedure. Ruchkin and Glaser (1978) describe simple digital filters and present their characteristics. More generally, Cook (1981) has developed a Fortran program, based on the methods of Ackroyd (1973), which determines the optimal values for a set of weights for a non-recursive filter, given sampling interval, bandwidth, and number of weights desired. Glaser and Ruchkin (1976) present a mathematical discussion of digital filters oriented to the psychophysiologist.

A more elaborate digital method for filtering voltage x time function is Wiener filtering (Walter, 1968; Wiener, 1964). Naitoh and Sunderman (1978) outline the application of this method to ERP data. As they describe it, an estimate of the frequency characteristics of background noise is made from a comparison of the spectra of the average ERP with the average of the spectra of single-trial ERPs, the spectra being obtained via Fourier analysis. This noise estimate is then used to correct the single-trial spectra. Finally, the original ERPs are regenerated via inverse Fourier transforms of the corrected spectra. Naitoh and Sunderman review evidence that Wiener filtering does not adequately preserve high-frequency information. Furthermore, they suggest that as a technique for general use the slight improvement in signal-to-noise ratio is not worth the trouble (see also Carlton & Katz, 1980; Ungar & Basar, 1976). However, they describe special circumstances for which it might be very appropriate.

Other than for simple smoothing, digital filters are perhaps most commonly employed prior to a pattern recognition procedure such as Woody filtering (see Section 3.2.3.3). However, they are potentially appropriate for any voltage x time function. They deserve serious consideration in the

laboratory, particularly given the continually decreasing cost of additional computation.

3.3 Data Reduction Techniques

3.3.1 Introduction

Although the headings "Signal Extraction Techniques" and "Data Reduction Techniques" serve to illustrate the fact that these are distinct processes to which psychophysiological signals are subjected, they are not mutually exclusive. One technique included under the heading of Signal Extraction procedures which would also fit under the present heading is linear Stepwise Discriminant Analysis. Discriminant analysis techniques serve both to provide a method of signal extraction and pattern recognition and, at the same time, to reduce the magnitude of the data set to a much smaller subset of variables. Another technique, Principal Components Analysis (PCA), which will be discussed under the present heading, could have been included in the Signal Extraction section. As with Discriminant Analysis, the PCA procedure serves to reduce the size of the data base from numerous dimensions to a relatively few "components". In addition, the PCA technique does not require the restrictive, a priori assumptions of group membership which characterize the discriminant analysis procedure. Thus, we do not wish to assert that any of the techniques illustrated in this chapter fit into a single category but instead that there are distinct stages in the process of data analysis.

A major problem in the analysis and interpretation of psychophysiological data is the determination of the specific criteria by which a signal is defined. For example, if one averages single-trial data, one makes certain assumptions about the signal and noise distributions that

underly the data: that portion of the voltage x time function which is temporally invariant over repeated presentations of a stimulus is defined as the "signal", while other, randomly varying portions of the epoch which are reduced as a result of averaging are defined as the "noise". Even if we adopt the signal/noise model implied by the averaging procedure, the problem of determining the important features of this "signal" remain. One commonly employed procedure for subdividing the average signal is to define its features on the basis of their relationship to the experimentally induced variance. In this case, the important features of the signal become identical with the components of variance in the data set. This type of definition of features or components of the voltage x time function requires not only the proper use of signal extraction and data reduction techniques but also the exercise of tight experimental control. Since components of the signal are defined in terms of their relation to experimentally manipulated variance, poor experimental design can lead to spurious components. Thus, another point to be emphasized is that the proper use of methods of analysis can provide the investigator with useful information only within the framework of good experimental design.

As mentioned above, the signal extracted from the raw or average voltage x time function is typically subdivided into features or components which are related to the experimentally induced variance. Each of these derived components can be thought of as a linear combination of weighting coefficients x time points. The problem with this approach, however, lies in the fact that there are an infinite number of possible linear representations for a vector of voltage x time values. Therefore, criteria must be adopted to aid in the selection of a subset of possible linear combinations. The determination of these weighting coefficients and their

application to the voltage x time signal will be the primary topic of the remainder of this section.

3.3.2 Peak Measurement

The identification of a peak in a voltage x time vector is perhaps one of the oldest measurement procedures in psychophysiology. The procedure is relatively simple, and it provides both amplitude and latency information. Although there are several methods of defining the peak of a component, they all involve a simple linear combination rule which assumes a weighting coefficient for each time point. This rule typically involves setting all of the weighting coefficients to zero, except for the one weighting coefficient a(x) which corresponds to the time point t(x) at which either the largest or smallest voltage is observed within a prespecified temporal window. This coefficient is set to one. Thus, in the case of peak measurement, the component derived from the voltage x time vector is defined as a single point. The principal advantages of this measurement procedure are its intuitive appeal and computational simplicity. Peak measurement algorithms represent a direct analog of the visual inspection of voltage x time data, with the added advantage of an easily standardized selection procedure.

A few representative procedures for peak measurement will be presented. The identification of "peaks" (single-point events) as zero crossings along the voltage x time function was proposed in the mid-1960's (Ertl, 1965; Ertl and Schafer, 1969). The method was suggested for peak identification in average ERPs and provides a reliable means for determining latency information. However, amplitude information is not available, since the peak has been defined as the zero point. Another method of peak

identification which has been widely employed involves selecting either the largest or smallest voltage within a prespecified temporal window and defining it as the peak. Amplitude information can be obtained from a base-to-peak difference, with the baseline usually being defined as some relatively inactive portion of the voltage x time function, such as that for some period prior to stimulus presentation. Alternatively, a peak-to-peak difference can be derived. In both cases, latency information is provided by the time point t(x) at which the largest or smallest voltage is obtained. The peak or peaks of the voltage x time function can also be defined in terms of the intersection of the tangents of their positive and negative slopes. Amplitude and latency information is also provided by this method.

All of the peak measurement techniques outlined above provide latency information, and, with the exception of the zero crossing technique, all give amplitude information. Note that each of the procedures makes the assumption that the psychophysiological component of interest can be defined as a single point in the voltage x time vector. When measuring a well delineated peak with a large signal to noise ratio (either the single trial or average), this assumption would appear to be appropriate. Examples of psychophysiological signals which would meet these criteria include the cardiac R-wave, the skin conductance response, and the systolic and diastolic peaks in the blood pressure cycle. However, even peaks which normally are sharply defined can easily become obscured by non-systematic variance, producing spurious measurements. Another disadvantage of defining a component in terms of a single point is the loss of information concerning the morphology of the voltage x time function. This information, which may be of benefit to the investigator, is discarded prior to analysis of the peak measurement. In effect, all information other than a single point in

the voltage x time function is defined as noise in the peak measurement procedure.

Other signal extraction and data reduction techniques also make assumptions about the nature of the signal and noise distributions. However, a subset of these provide information that is similar to that given by the peak measurement techniques while also retaining some morphological information. For example, the polarity histogram is one measurement technique which provides amplitude information in the form of probability instead of voltage (Callaway and Halliday, 1973; Kubayashi and Yaguchi, 1981). The procedure is performed by incrementing a frequency count whenever an individual time point in the voltage x time function is above or below a zero baseline. A component is then defined whenever the time x probability histogram exceeds some criterion value. The advantages of the technique include its computational simplicity and relative insensitivity to random fluctuations in the voltage x time signal. Some morphological information is also retained in the form of probability values.

Another procedure which provides amplitude as well as morphological information (symmetry and peakedness) has been proposed by Callaway, Halliday, and Herning (1983). In this procedure, called PEAK, a grand average template is computed. The important features (peaks and troughs) of the template are defined by means of a standard algorithm. Lagged correlations are then computed between the template and the individual voltage x time vectors. Components in the voltage x time vectors are defined as the maximum lagged correlations between the template and the individual vectors. A series of measurements are then made on the features such as amplitude, latency, peakedness and symmetry. Other component measurement techniques such as area measurement and PCA, which also provide

alternatives to traditional peak measurement procedures, will be discussed below.

Another disadvantage of the peak measurement methodology is the difficulty encountered in defining the peak of a relatively slow component. Can a single time point accurately represent a slow component--and, even if it could, which point would be selected? Several psychophysiological signals would qualify as slow components (e.g., respiration, skin conductance response, contingent negative variation-CNV). Techniques such as area measurement and PCA may provide a more appropriate representation of these components.

In addition to the limitations mentioned above, peak measurement techniques also fail to provide information concerning component overlap. The measurement of a single point does not permit the assessment of the actual number of temporally overlapping components which may jointly be responsible for the voltage recorded at the specific time point. Several examples of this particular problem have been addressed in the ERP literature (Donchin, Tueting, Ritter, Kutas, & Heffley, 1975; Squires, Donchin, Herning, & McCarthy, 1977). While carefully designed factorial experiments can alleviate this problem to some degree, a better solution lies in the application of a procedure which will permit a direct evaluation of the overlapping components.

A final problem concerns independence among peaks when several peaks are measured in the voltage x time function. This is particularly important if statistical inference techniques are to be applied to the data, since most of these techniques assume independence among measures.

In summary, peak measurement procedures must be applied with caution when they are used to define a psychophysiological component. It must be realized that data reduction may result in the loss or distortion of relevant information. However, the techniques outlined above can provide useful information in situations in which the psychophysiological signals are relatively fast, are well delineated, and possess a high signal-to-noise ratio.

3.3.3 Area Measurement

Like peak mesurement, the measurement of the area of a psychophysiological component can also be conceptualized in terms of a linear combination of time points. In this case, however, the weighting coefficients which correspond to the temporal region of the component are set to one while the rest of the weighting coefficients are set to zero. Thus, unlike the peak measurement procedures, area measurement defines the component of interest in terms of a range of contiguous time points. These points are then integrated relative to a baseline to produce the area measurement of the component. The assumption underlying the use of area measurement is that the psychophysiological component is most accurately represented by the area of some specific epoch along the voltage x time function. This appears most reasonable in the case of slow components such as the skin conductance response, respiration, and CNV.

The measurement of the area or amplitude of a component is performed relative to some baseline. In most cases the baseline is defined as that portion of the voltage x time vector which occurs prior to stimulus presentation. It is assumed that the baseline represents an inactive portion of the vector. However, this is not always the case. In some

situations, anticipatory activity is present (e.g., CNV). In this case another method of defining an inactive baseline is required. One such method is the use of "trimmed" averages which are relatively insensitive to extreme deviations in the data (see Donchin and Heffley, 1978).

Like peak identification, area measurement also possesses a good deal of face validity, since many psychophysiological signals extend over more than a few time points. The need for elaborate computational algorithms is also minimized by area measurement. Furthermore, area measurements are less susceptible to modest amounts of latency jitter in the component, as well as less sensitive to random amplitude variations in a few time points, than is peak measurement. The degree of insensitivity to random fluctuations is a function of both the number of points included in the area and the temporal range of the latency variability.

Although area measurement presents a distinct advantage over peak identification in some cases, it still fails to deal adequately with several measurement issues. The determination of integration limits is often difficult and/or arbitrary due to the poor resolution of component limits in the raw or average voltage x time function. The issue of the establishment of reliable integration limits becomes less of a problem with components which are easily recognized. The issue of component overlap is also not addressed by the area measurement procedures: It is difficult to assess the relative contribution of overlapping components to the voltage measured at either one or several time points. As has been mentioned above, one way to lessen this problem is to control the experimental variables which are known to affect the amplitude and latency of the overlapping components. Finally, as with peak measurement, area measurement techniques may fail to provide

the investigator with a clear, detailed picture of the morphology of the voltage x time function.

In summary, although area measurement procedures alleviate some of the problems encountered with peak identification techniques, there still remain unresolved issues. Area measurement would appear to be most appropriate when non-overlapping, slow components are evaluated.

3.3.4 Principal Components Analysis (PCA)

3.3.4.1 Introduction

Unlike discriminant analysis, PCA does not require that the subclasses be known *a priori*. Thus PCA makes less restrictive assumptions about the number of relevant categories into which the data will be subdivided. This is particularly useful to the investigator when the nature and number of subclasses is unknown prior to the analysis. In addition to the pattern recognition information garnered from PCA, the technique also provides a means by which a huge data base is reduced to a few components which most parsimoniously describe the experimental variance. Although the PCA procedure has been employed most frequently in the analysis of ERP data, it is clearly relevant to the analysis of other psychophysiological signals.

Like peak and area measurement techniques, PCA can also be conceptualized in terms of a linear combination of time points. To reiterate, the peak measurement procedure defines the psychophysiological component as a single time point in the voltage x time function. The other time points are discarded prior to analysis. In the case of the area measurement, the psychophysiological component is defined as the integration of equally weighted values at several time points. Area measurement

represents a distinct improvement over peak measurement procedures in the assessment of slow components. However, neither procedure addresses the issues of the selection of optimal weighting coefficients or the effects of component overlap on the observed voltages.

Unlike the peak and area measurement techniques, the PCA procedure employs the complete voltage x time data matrix to determine the weighting coefficents. In the present case we will be describing the R-PCA procedure, which involves the computation of a time point x time point input matrix. Other investigators (John, Ruchkin, & Villegas, 1964; John, Ruchkin, & Vidal, 1978) have suggested the usefulness of the Q-PCA procedure, which involves the computation of a waveform x waveform input matrix. In the former case the interest is in the relationship among time points across the voltage x time function. In the latter case, the analysis provides information concerning the relationship among individual waveforms in the data matrix. Although the present discussion will be concerned with the R-PCA procedure, its general points also apply to the Q-PCA technique.

In terms of providing optimal weighting coefficients for the determination of components, PCA is clearly preferable to the methods employed in peak and area measurement. In the case of the PCA procedure, the weighting coefficients (component loadings) represent the contribution of the derived component to the variance at each time point in the voltage x time function. Another advantage of the component extraction procedure employed in PCA is that the weighting coefficients associated with each component are uncorrelated with the weighting coefficients associated with each of the other components. Thus, the component scores computed from the linear combination of time points x weighting coefficients are orthogonal. Therefore, in contrast to peak and area measurements, PCA permits the

investigator to assess the independent effects of the experimental manipulations on temporally overlapping components (Donchin, et al, 1975; Glaser and Ruchkin, 1976).

As with the peak and area measurement procedures, the method of determining the weighting coefficients in the PCA procedure implies a particular definition of the psychophysiological component. The PCA procedure defines a component in terms of the covariation between time points in the voltage x time function. A pattern of high covariation among time points implies that a specific component (source of variance) can be assumed to be influencing them jointly. These derived components are represented in terms of the variance in the data. The component score produced by the linear combination of the time points x weighting coefficients provides a measure of the magnitude of a specific component in a specific voltage x time function. Thus, for each PCA component a separate weighting coefficient is obtained for each of the time points, and a separate component score is derived for each voltage x time function in the data matrix. An example of a component loading plot is presented in Figure 4. This figure displays four sets of component loadings and the grand mean waveform from an ERP experiment. There are 128 component loadings which correspond to the 128 time points in the waveform. A separate set of loadings is calculated for each of the four components.

Insert Figure 4 About Here

There are several assumptions that underlie the PCA model. It is a linear model and thus assumes that the derived components simply sum together to produce the voltage x time function without interaction. A

second assumption concerns the sources of variability in the data. It is assumed that the sources of variance in the data are orthogonal. Although there is no foolproof method of assuring that this assumption is met, good experimental design in terms of the factorial manipulation of experimental variables which are believed to influence the major sources of variance is one way to minimize intercomponent correlation (Donchin and Heffley, 1978). Techniques are also available for testing the assumption of orthogonality (Harman, 1967). In cases in which two or more sources of variance are highly correlated across voltage x time functions, PCA will yield a set of weighting coefficients and a single component score which represent a composite of these correlated components. The interpretation of this composite component in terms of the voltage x time data set will be misleading (Roessler & Manzy, 1981; Wastell, 1981a). A third assumption of the PCA model concerns the domain of component variability. PCA can reliably and efficiently handle variability in the amplitude of the component. On the other hand, variability in the latency of the component over voltage x time functions can cause substantial problems in the interpretation of the derived components. The PCA procedure does not discriminate between variance in the data which is due to variations in amplitude of the underlying component and that due to variations in latency of the underlying component. Therefore, if both the amplitude and latency of a component are changing over trials, PCA will not be able to distinguish the two dimensions. In the case of latency variability, some attempt needs to be made to decrease the variability over trials prior to the use of the PCA technique (Picton and Stuss, 1980). One such procedure which is described in the present chapter (see Section 3.2.3.3) is the adaptive filter for the analysis of variable-latency neuroelectric signals (Woody, 1967; Woody and

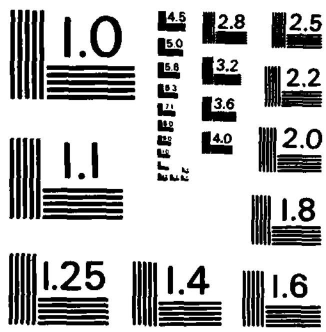
Nahvi, 1973; Navhi et al, 1975). Callaway, et al (1983) demonstrate the improvement in PCA results which latency correction can provide.

3.3.4.2 Appropriate Experimental Design

The assumptions of the PCA model which were outlined above and the ease with which they can be violated, suggest that PCA cannot be blindly employed in the analysis of psychophysiological data. The exercise of good experimental design as well as sensitivity to the assumptions of the PCA model are of paramount importance if the technique is to provide valid information. The PCA technique represents a multi-step procedure for the analysis and interpretation of psychophysiological data. Each step in the procedure requires forethought about the assumptions of the model and the design of the experiment. The initial step, and perhaps the most important, concerns the design of the experiment. There are several issues which must be considered prior to the design of an experiment destined for PCA.

The first issue concerns the second assumption of the PCA model mentioned above, that the major sources of variance are orthogonal. One method to minimize intercomponent correlation is the factorial manipulation of the major sources of variance. A second issue to be considered during the design of the experiment is that the number of cases (typically, subjects x conditions) should exceed by a factor of 10 or more the number of variables (typically, number of time points) in the voltage x time function (Picton and Stuss, 1981). As the number of cases decreases relative to the number of variables, the stability of the component structure will also decrease. In terms of a practical example, this means that a voltage x time vector with 60 time points (variables) would require 600 separate cases to insure stability. A third issue to be considered during the design of the

AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF 2/9
INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN
COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL.
UNCLASSIFIED 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

experiment concerns the requirement of the PCA model that the number of variables be of sufficient quantity to determine a stable component structure. For many psychophysiological data sets, this mathematical precondition is usually not a problem since a large number of variables produce relatively high loadings on each component. However, if underdetermination of the component structure is suspected (too few variables having high loadings on a specific component) there are several techniques which permit the investigator to assess the resulting instability (Thurstone, 1935; Mulaik, 1972; Tucker, Note 1).

3.3.4.3 Selection and Computation of the Input Matrix

Once the investigator has designed the experiment and collected the data, the next step is to decide on the type of input matrix to be employed in the PCA solution. This selection has important implications for the interpretation of the resulting component structure. The input matrices which are accepted by most PCA programs include mean crossproducts, covariance, and correlation matrices.

Calculation of the mean crossproducts matrix involves summing the products of crossmultiplication of the voltage values for all time points with all other time points. Note that in this case all of the experimental variance is analyzed since neither the mean nor the variance at each time point is removed from the data set before analysis. The fact that the mean is not subtracted from the crossproducts matrix has certain implications for the component structure. To begin with, the loadings of the first derived component usually duplicate the grand average voltage x time function. Second, large base-to-peak deflections in the voltage x time function will produce components even when they are not influenced by the experimental

manipulations (Donchin and Heffley, 1978). The use of the crossproducts matrix would appear to be most appropriate when the investigator wishes to retain information about the absolute variations in amplitude as well as the polarity of the corresponding component in the raw data (see Ruchkin, Sutton, & Stega, 1980; Squires et al, 1977).

The calculation of the covariance matrix is similar to that of the crossproducts matrix, with the exception that the grand average voltage x time vector is subtracted from the individual voltage x time functions prior to the computation of the crossproducts. Thus, the portion of variance which is contributed by the differences between the variable (time point) means is removed in the process of calculating the covariance matrix. In terms of the component structure which will be derived from the covariance matrix, the important issue will be the degree to which the individual voltage x time functions differ from the grand average, not the absolute amplitude or polarity as was the case for the crossproducts matrix. Thus, component scores will reveal relative rather than absolute differences in the component. The covariance matrix has been used most frequently in the analysis of ERPs, since the differences among ERPs relative to the grand mean waveform are usually of primary importance (see Isreal, Chesney, Wickens, & Donchin, 1980; Ruchkin et al, 1980).

The correlation matrix is another option in the selection of the input matrices for the extraction of principal components. The calculation of the correlation matrix requires that the mean of each variable be subtracted (as in covariance) and additionally that the difference be divided by the variable's standard deviation. Thus, in the case of the correlation matrix the variance attributed to the differences between the time-point means as well as the variance due to differences in time-point variability is removed

during the process of calculation of the matrix. Essentially, each time point value is converted prior to PCA to a standard z-score, based on that point's mean and variance across voltage x time functions. The components extracted from the correlation matrix will be similar to those derived from the covariance matrix, with the exception that the loadings will be more uniform across the length of the component due to the standardization of the variables (Donchin & Heffley, 1978). Thus, in the case of the correlation matrix, the loadings will not reflect the component morphology as well as when the covariance matrix is employed. This standardization also serves to obscure the magnitude of differences in variance across time points. This may result in the assignment of relatively high loadings to time points at which differences are small.

As can be seen from the previous discussion, the choice of an input matrix for the PCA procedure constrains the conclusions which can be drawn from the derived component structure. Thus, the investigator must take a careful look at the specific questions which are to be addressed with the PCA, prior to the selection of the input matrix.

3.3.4.4 Extraction of Principal Components

The third step in the PCA process involves the extraction of the weighting coefficients to be used in the linear combination of time points. The extraction procedure, consisting of a sequence of standard matrix manipulations normally performed by packaged statistical software, produces one vector of weighting coefficients for each of the derived components. A separate weighting coefficient is derived for each of the time points in the voltage x time function. Thus, if six components are extracted from a series of voltage x time functions, each composed of 60

time points, there would be six sets of 60 weighting coefficients derived in the PCA procedure. As has been mentioned above, a vector of weighting coefficients represents the contribution of the derived component to the variance at each time point in the voltage x time function. The weighting coefficients associated with each component are uncorrelated with the weighting coefficients associated with each of the other components.

The orthogonality of the components produced by the PCA technique represents a distinct advantage over the peak and area measurement procedures in terms of later inference testing. Univariate analyses of variance (ANOVA) can be performed on the component scores for each of the components. On the other hand, computation of separate ANOVAs for each peak or area measurement is of doubtful validity due to the possible correlation between measures in different parts of the voltage x time function.

The first component extracted in the PCA accounts for the largest proportion of systematic variance in the data matrix. The second derived component accounts for the largest possible percentage of residual variance and is orthogonal to the first component. This process of component extraction continues until all possible components have been derived.

It must be noted that the components derived via PCA need not reflect the physiological generators underlying the recorded voltage changes in a one-to-one fashion. Instead, the components represent merely one summary of the systematic variance present in the data. Theoretical inferences and converging measurement operations are required to verify the relationship of PCA components and physiological components.

One of the goals of the PCA technique is the reduction of the data base to a subset of meaningful components. That is, the hope is that a few orthogonal dimensions (components) will be able to account for most of the

variability in the raw data, or that most of the information in the raw data can be more simply represented. Intuitively, this is possible to the extent that the original observation time points are redundant. Determination of the number of components to retain is usually based on criteria such as the amount of variance accounted for and the parsimony of interpretation of the component structure. Several statistical methods have been suggested to assess the number of components to retain (Cattell, 1966; Humphreys and Montanelli, 1975; Kaiser, 1960; Montanelli and Humphreys, 1976; Tucker, 1973).

One point which is specifically relevant to component extraction with psychophysiological data concerns the temporal range of the components in the voltage x time function (Wastell, 1981b). The PCA procedure initially selects components associated with relatively slowly varying regions of the voltage x time function since these components typically encompass a large amount of the variance. Somewhat faster components such as the P300 are then selected. Components which extend over a relatively limited temporal range will be extracted much later in the PCA procedure. Therefore, by virtue of the component extraction procedure employed in PCA, some fast components will not constitute a sufficient amount of variance to produce a component which will meet the selection criteria. This point is especially important if the voltage x time functions consist of both slowly and quickly varying components.

3.3.4.5 Rotation of Component Loadings

Once the desired number of components has been extracted from the input matrix, the next step usually involves trying to simplify the component structure. In most cases the component loadings for each derived

component vary across the entire voltage x time function, because of a non-zero correlation among time points. The purpose of the rotation procedure is to simplify the pattern of loadings so as to localize each component to a portion of the voltage x time function.

The Varimax rotation procedure has been frequently used with ERP data and provides one method by which the interpretability of the component structure can be enhanced. The procedure retains an orthogonal component space while maximizing the variance of the component loadings by attempting to drive the high loadings to unity and the low loadings to zero. Thus, the Varimax rotation maximizes the association between each component and a few time points and minimizes the association at all other time points for each component. The rotation redistributes the component variance among the time points but does not alter the goodness of fit of the component model. Note that the PCA extraction procedure provides the component structure while the rotation temporally localizes the components, thereby permitting the evaluation of the components in terms of the original voltage x time functions. In terms of psychophysiological data, the Varimax procedure emphasizes the peak of the signal and is therefore analogous to a base to peak measurement.

Following the completion of the rotation procedure, the next step is to compute the linear combination of time points x weighting coefficients for each component. This transformation will produce a separate component score for each voltage x time function in the input matrix. The component score represents a measure of the magnitude of a specific component in a specific voltage x time function.

3.3.4.6 Inference Testing

The majority of psychometric studies which employ the PCA technique terminate prior to the calculation of component scores. In many cases investigators are solely interested in the association between the principal components and the observed data. Component loadings are sufficient to provide this information. The psychophysiologist, on the other hand, is also concerned with the effect of experimental manipulations on the components derived from the voltage x time data. In this case the component scores as well as the loadings are of interest. Calculation of the component scores permits the investigator to locate the observed voltage x time functions in a simpler and presumably more meaningful component space. Differences among the component scores reflect the effect of experimental manipulations on the principal components and may be subjected to inference testing procedures (see Section 5).

3.3.5 Summary Comparison of Data Reduction Techniques

At this point it is appropriate to summarize the advantages and disadvantages of PCA in the analysis of psychophysiological data, relative to simpler data reduction techniques. To begin with, PCA provides an objective and statistically based method for identifying and computing linear combinations of time points x weighting coefficients. This serves to reduce experimenter bias in the selection and definition of the psychophysiological components. Another advantage of the PCA procedure concerns the method of calculating the weighting coefficients. In the peak and area measurement procedures, weighting coefficients are set to either zero or one. The PCA technique permits the assignment of graded weighting coefficients on the basis of the contribution of the derived component to

the variance at each time point in the voltage x time function. Thus, the entire data set is employed in the calculation of component scores, rather than a few time points. This serves to increase the sensitivity of the experimental procedures as it attenuates the effects of noise and sampling fluctuations on the components. Furthermore, unlike the peak and area measurement techniques, PCA provides information about both the amplitude variability and the morphology of the voltage x time functions. Amplitude information is available in the form of component scores. The morphological characteristics of the component are provided by the weighting coefficients. PCA also gives the investigator information about the degree of component overlap, provided the underlying components are not highly correlated. Since the components derived from the PCA are orthogonal, univariate tests of significance may be appropriately applied to the component scores. Finally, PCA provides an efficient summary of a very large data base by providing a simpler and therefore more interpretable data structure.

Although PCA presents numerous advantages over some traditionally employed psychophysiological analysis techniques, there are some limitations which should be mentioned. For example, the PCA model assumes that the components embedded in the voltage x time function are temporally invariant over trials. In cases in which this assumption is not met, PCA confounds the amplitude and latency variability of the components and provides a component structure which is difficult to interpret. There are, however, several techniques which can be employed as preprocessors to reduce the latency variability prior to employing the PCA technique (e.g., Woody filter or other autocorrelation measures). The transformation process employed in the PCA is certainly not as intuitively clear as that used in peak or area measurements. This may sometimes lead to confusion when raw voltage x time

functions are compared with the reduced component structure. Another point to consider is that components in the voltage x time functions which span a relatively few time points may not constitute sufficient variance to meet the component selection criteria prior to rotation. Finally, since PCA does in fact employ the entire time point x time point data matrix, substantial computing power is required to carry out the transformations.

3.4 Spatial Analysis

3.4.1 Introduction

Although some psychophysiological signals can be treated as reflecting the activity of a single structure (as in the case of heart rate), in other cases the signal reflects the activity of what are functionally multiple generators (for example, EEG). Furthermore, the signal produced by these generators, propagated through space to the body surface, can vary as a function of the spatial characteristics of the generators and the conductivity characteristics of the structures interposed between the generators and the skin. As a result, the signal recorded at the surface will depend on the location of the electrode or other transducer.

In some cases the variability due to electrode location is not of interest to the psychophysiologist but constitutes merely a source of error to be eliminated. For example, EKG morphology depends greatly on electrode location. However, the psychophysiologist might only be interested in interbeat interval. Thus, variation in the morphology of the EKG waveform with electrode position can be ignored. Of course, when variation due to location is ignored in this way, the psychophysiologist is assuming that a single "channel" or "generator" is of interest and that the variability

observed at different locations on the body surface is irrelevant. This model is more often adopted for measures of autonomic activity than for EEG or EMG. However, by ignoring the spatial distribution over the body surface of the psychophysiological signal we may miss a relevant part of the information provided by the signal.

Although there are serious problems in making inferences about location of the ERP component generators from the scalp distribution, measures derived from multi-electrode recordings can still be very useful as an empirical method for defining components (see Donchin, 1978). In fact, if a component recorded at the scalp represents the sum of many fields generated by the activity of neurons functionally linked together (although not necessarily localized in a specific brain structure), the scalp distribution of a component will reflect its spatial properties. If we accept this basic model, and if we use scalp topographic information merely to infer functional, not physical, generators, the actual relationship between anatomical generators and their scalp manifestations need not be known. To this end, it is only necessary to record from those locations that allow us to discriminate among functional systems.

The remainder of this section will be concerned with a brief description of some procedures devised to study the spatial distribution of psychophysiological measures. Although these procedures have been devised for analyzing event-related brain potentials, they can be applied to any other measure that can be recorded simultaneously from multiple locations.

3.4.2 Isopotential maps

Isopotential maps are one way of expressing the values of a psychophysiological variable at different locations on the body surface.

They involve recording at a large number of locations in order to obtain an accurate description of the similarities in voltage between different points of the body surface at a particular time point. Isopotential maps have been most frequently used for the EEG.

In an isopotential map (e.g., Ragot & Remond, 1978), the body surface is schematically represented on paper in the same way that terrain is represented in topographic maps. Voltage values observed at any location on the body are presented at the corresponding points of the map. Values of the intervening points are extrapolated by means of algorithms that typically rely on values at adjacent points. Points with equal values are then connected by lines, and a convention is adopted to distinguish positive and negative values.

Isopotential maps constitute only a graphical representation of the data, and do not therefore imply any particular assumptions (beyond those concerning interpolation). However, they do not simplify the structure of the data, and therefore they do not qualify as signal extraction techniques. Rather, they are a preliminary tool for investigating the spatial distribution of the psychophysiological variable, where no assumptions about signal and noise are made.

A particular kind of isopotential map is the spatiotemporal map (Remond, 1962). In this map, one of the axes is given by time. Therefore, the spatial information is restricted to a line, but information about the variation over time of the spatial distribution is included. As above, this kind of map is more a data description technique than a signal extraction procedure. The problem of defining the signal remains unsolved.

Another kind of spatial map is the Significant Probability Map (Duffy, Bartels, & Burchfiel, 1981). This kind of map plots "z" or "t" statistics

obtained by the comparison of pairs of values from two data sets. Maps for different time points are compared. A signal is defined as "those aspects of the distribution which differentiate significantly between two sets of data". Note that this kind of definition yields a signal that is specific to the data sets used, and comparisons between data obtained in different experiments are problematic.

3.4.3 Univariate and Multivariate Approaches to Spatial Analysis

An isopotential map is essentially a graphical way of representing the information obtained with a multiple-electrode recording. Because it does not make any distinction between signal and noise, it does not qualify as a signal extraction technique. However, signal extraction from an isopotential map can be accomplished in at least two ways. A peak detection algorithm (see Section 3.3.2) can define a signal. Alternatively, the signal may be defined on the basis of a pattern in point-by-point t-tests between what is understood to be signal-present and signal-absent conditions (see Significant Probability Mapping in Section 3.4.2).

However, the use of typical univariate techniques to test inferences about data from multiple electrode recordings is unsatisfactory for two reasons. First, the large number of resulting significance tests greatly inflates experiment-wise error rate. Standard adjustment of the alpha level is likely to undercorrect for this problem because error variance is likely to be correlated across recording sites. Second, univariate analysis provides little information about effects or patterning at different sites. A further limitation of standard methods of signal extraction and inference testing with isopotential maps is the inability to distinguish between overlapping sources of activity at each time point.

other at each specific frequency. Since physiological processes are not perfect sinusoids, but occur over a band of frequencies, we have developed a summary statistic that describes the proportion of shared variance between two systems over a band of frequencies (see Porges, Bohrer, Cheung, Drasgow, McCabe, & Keren, 1980). We have labeled this statistic the weighted coherence (Cw). In our laboratory, Cw has been used primarily to describe the relationship between heart period and respiration. However, the application of Cw is not limited to the assessment of the coupling between respiration and heart period activity, but may also be used to determine the proportion of shared variance between any two processes that fit the statistical assumptions for spectral analysis.

Spectral analysis is based upon a model that assumes that the constituent periodic components of a time series are statistically "independent" and linearly additive. There are situations in which one frequency component in a system could trigger a faster frequency. For example, consider a physiological system in which four breaths occurred before there was a general shift in blood pressure. Both frequencies would be manifested in the spectrum of blood pressure. Using traditional spectral analysis, one would assume that the periodic components were independent. However, by using a spectral technology called "polyspectral" (see Brillinger, 1975), it is possible to identify potential "coherences" between two frequency components within one physiological process or between two different frequency components represented in two different physiological processes.

into a set of pure sine wave of different frequencies, with a particular amplitude and phase angle for each frequency.)

4.2.3 Other Frequency-Domain Methods

There are other frequency-domain techniques. A simple and often visually appealing method is "zero-crossing". This method quantifies the frequency with which a waveform crosses an arbitrary baseline. It provides a relatively accurate estimate of the frequency of the process if, and only if, the process contains only one periodic component and is not contaminated by background noise. The periodogram is effective at finding periodic components and may be efficiently calculated using the Fast Fourier Transform. However, the periodogram has poor statistical characteristics and should not be used without appropriate frequency-domain smoothing (see Bohrer & Porges, 1982).

Periodic covariation may be described with cross-spectral analysis. Cross-spectral analysis generates a coherence function which is a measure of the best linear association of each observed rhythm in one variable with the same rhythm in a second variable. The coherence is the square of the correlation between the sinusoidal components of the two processes at a specific frequency. The coherence at any specific frequency is the square of the cross-spectral density divided by the product of the spectral densities of each series at the specified frequency. Note the similarity of this equation with the calculation of a squared correlation coefficient; the cross-spectral density parallels the squared cross-products and the spectral density parallels the variances. Conceptually, the coherence may be thought of as a time-series analog of omega squared (see Hays, 1981) or the proportion of variance accounted for by the influence of one series on the

response which occurred outside the confidence intervals of the forecasted values. Autoregression models may be as simple as a linear forecast (i.e., projecting best linear fit from the baseline) or may involve higher order models. Individuals interested in applying time domain forecasting and prediction models to detect the impact of an intervention are encouraged to study Box and Jenkins models (Box and Jenkins, 1976) and to be familar with the interrupted time-series model described by Campbell and Stanley (1966).

If the goal is to describe a periodic signal which represents only a small percentage of the total variance of the series, then the successful application of time domain techniques will be limited to the experimenter's ability to filter the data by removing trend and periodicities other than the one of interest (see Sections 2.2.3 and 3.2.4). This requires a priori knowledge of the underlying periodic structure of the process.

In contrast to time domain techniques, frequency domain techniques are those based upon the spectral density function, which describes how the periodic variation in a time series may be accounted for by cyclic components at different frequencies. The procedure estimates the spectral densities at various frequencies and is called "spectral analysis". For bivariate series, the "cross-spectral" density function measures the covariances between the two series at different frequencies.

Spectral technology decomposes the variance of a time series into constituent frequencies or periodicities. There is a mathematical relationship between the time domain correlation procedures and spectral analysis. The spectral density function is the Fourier transform of the autocovariance (unstandardized correlation) function, and the cross-spectral density function is the Fourier transform of the cross-covariance function. (The Fourier transform is an algebraic method of decomposing any time seri

time-shifted version of itself. If the time series is periodic, the plot of the autocorrelations (the autocorrelogram) at different time lags will be periodic. Similarly, a cross-correlation is the correlation of one time series with a time-shifted version of a second time series. The cross-correlation function provides information regarding the statistical dependence of one series on another. If the two time series are identical, the peak value of the cross-correlation function will be unity at the lag that makes the two series identical and less than unity at all other lags. In most cases, since the second series is not only a time-shifted version of the first series, the peak value of the cross-correlation will be less than unity.

Autocorrelation techniques are effective in detecting periodicities only when the series are characterized by a relatively pure sinusoid, uncontaminated by other influences. Cross-correlation techniques lose their effectiveness and sensitivity to assess the communality between two series when the difference between the series is more than a temporal displacement.

Autoregression techniques are more commonly used in developing models of baseline activity and using the model to forecast into the future. These techniques consist of predicting the value of a time series function at a particular time on the basis of previous values of that function. In a multiple regression sense, each previous time point serves as an independent predictor variable to which a weight is assigned. Stock market forecasting "systems" are dependent upon this type of modeling. Once the model is generated, confidence intervals can be calculated for the forecasted values. In the case of psychophysiological research, one could define a significant response in any physiological system, on any trial, for any subject, by evaluating whether the stimulus manipulation produced a physiological

describe the periodic characteristics of spontaneous EEG and fits nicely into our conceptualization of rhythmic generators in the central nervous system.

4.2 Time Series Analysis: Definitions and Methods

4.2.1 The Definition of a Time Series

Although most psychophysiological data are presented in terms of mean levels within or across subjects, the sequential pattern, on which the mean is based, may contribute important information. Time-series statistics provide methods to describe and evaluate these patterns. A set of sequential observations, such as the circumference of the chest sampled every second or the time intervals between sequential heart beats, constitutes a time series. Mathematically, a time series may be described as a string of variables that are sequentially indexed, for example,

$$X_t, X_{t+1}, X_{t+2}, \ldots\ldots X_{t+n}.$$

In this example, the index t represents time.

4.2.2 Time-Domain and Frequency Domain Methods: An Overview

There are two basic approaches that may be used to describe and analyze a time series. The series may be represented and analyzed in the time domain or in the frequency domain. Time domain representations plot data as a function of time (see Section 3). Those time-domain methods which are most closely related to the frequency domain are based on autocorrelation and cross-correlation measures. As their names imply, the techniques are mathematical extensions of traditional correlational techniques. An autocorrelation is the correlation of one time series with a

The repeated-measures analysis of variance design tends to evaluate "pre", "during", and "post" stimulation periods. By partitioning the variance in this manner, the variance is divided into "treatment" or "time" (repeated measures) and error effects. The error tends to include the variance associated with individual differences among the subjects. To reduce the variance associated with individual differences (i.e., error variance in the analysis of variance design), potent manipulations are used. The objective of this strategy is to enhance the "signal" to "noise" ratio by maximizing the difference between the "baseline" spontaneous activity and the "response" elicited by the stimulus manipulations.

Ironically, massive treatments often violate the homogeneity of variance assumption of analysis variance. Although the analysis of variance is viewed as a "robust" test and is relatively insensitive to violations of the homogeneity of variance assumption in between-groups designs, in the repeated-measures design slight variations in the variance between repeated measures will produce difficulties in interpretation (see Section 5.2.3; Porges, 1979).

An alternative method of describing voltage X time functions is to incorporate "time-series statistics" into the experimental and quantitative strategies. Time series methods may be used to detect changes in the voltage X time functions in response to an event by describing the pattern of the function during baseline or stimulus conditions. Time series methods may be classified into two broad categories: Time domain and frequency domain. As a general rule all time series may be represented in either domain. However, certain data may be more easily or more appropriately described in one domain than the other. One domain may lead to a more natural interpretation. For example, the frequency domain is often used to

the experimental manipulation is conveyed by the mean level of the physiological process.

The second procedure is characterized by averaging across repeated trials. This procedure is based upon the view that physiological signals reflect neurally mediated responses to stimuli and are superimposed on the the background regulatory neurophysiological function. The averaging model assumes that the statistical distribution of the background activity is not influenced by the stimulus and that the background activity is identically and independently distributed. This means that the background activity is assumed to be distributed randomly. Thus, the two prevalent quantitative strategies of decomposing physiological response variance are insensitive both to the possibility that the "signal" is encoded in a parameter other than level and to the possibility that the "signal" is encoded in the background "noise".

Although all voltage X time functions are "time series," few psychophysiologists have used time series statistics (applied to the frequency domain) as analytic tools. Instead, most psychophysiological researchers have attempted to describe responses via more traditional descriptive statistics. For example, the experimental designs that have been prevalent in psychophysiology have involved traditional, repeated-measures, analyses of variance, which test effects of stimulus manipulations on the descriptive statistic of mean level. In a few instances, the pattern of the physiological response as a voltage x time function has been estimated with measures of variability such as the standard deviation (e.g., heart period variability). However, in most cases pattern is described by directional or polarity shifts in the voltage X time function (e.g., heart rate deceleration or P300).

4. Data Analysis in the Frequency Domain (by Stephen W. Porges)

4.1 The Description and Partitioning of Variance

The description of physiological activity as both dependent and independent variables is difficult. Physiological activity is seldom in a "binary" state which can be described as either being "on" or "off." Moreover, changes in level or frequency seldom are complete descriptors of physiological activity. The physiological systems of interest to psychophysiologists are continuously changing, reflecting the dynamic regulatory function of the nervous system. It would, of course, be naive to believe that these systems are sensitive solely to those variables we choose to manipulate in our experiments. Thus, we are faced with a series of paradoxical problems. For example, we may be interested in monitoring the central nervous system during manipulations of "mental effort" or "information processing." However, the dimensions of physiological activity which may be the most sensitive to the "neural mediation" of information processing may also be the most sensitive to the "neural mediation" of basic homeostatic function.

In the psychophysiological literature, two methodological procedures have been employed to deal with the problems of partitioning the impact of "stimulus processing" from the background "neurophysiological regulation." Both procedures reside within the time domain (i.e., the stimulus and the physiological activity are indexed by time). The first procedure is characterized by indexing the changes in mean level or variance produced by an experimental event. Implicit in this type of procedure is the notion of a "statistically significant" response. This notion is based upon a model which assumes that the variance of the physiological process associated with

contributions of as many components as the number of recording sites can be assessed. Note that, since the vector space is defined by the recording sites used, an appropriate choice of the electrode sites can greatly improve the resolution of the effects of overlapping components by the vector filter technique. However, components with different spatial distribution will be differently amplified, or filtered out, by a VF. On the other hand, the general approach of VA allows the investigator to distinguish between overlapping components. Procedures particularly devised to solve this problem are presented in Gratton et al (Note 2).

The lengths of the target vectors obtained at each time point can be directly submitted to standard inferential procedures. The practical result of VF is to "filter" the data for the component of interest (defined by the spatial distribution expressed in the target vector), with the filter output proportional to the goodness of fit between data vector and target vector. Therefore, VF qualifies as a signal extraction technique. A series of filter output values, one for each time point, constitutes a time series, whose values refer to the estimated contribution of the target component to each time point. This time series can be submitted to the analytical and inferential procedures described elsewhere in this chapter.

VF has several advantages in comparison with the traditional procedures of spatial analysis. First, it involves a small amount of computation. Second, it makes use of all the information available at any given time point. Third, it provides a tool for testing hypotheses concerning spatial distribution.

Since the distribution of the target component is established a priori, VF needs no cross-validation. Actually, VF itself can be considered as a test for the distribution of the target component. VF does not make use of the information obtained at different time points in determining the target component. While this can be in some cases disadvantageous (as the analysis is conducted separately for each time point), it is useful when the latency of the components is variable.

Another limitation is that VF is not able to distinguish between the overlapping contribution of several components to the same time point, unless their distributions correspond to orthogonal vectors in the vector space. In the latter case, for each time point, the independent

vector is postulated (i.e., at each time point, signal and error are uncorrelated). Therefore, the model adopted by VF assumes that the data vector at each time point is given by the sum of a target vector, with given orientation and unknown length, and of an error vector, with orientation orthogonal to the signal vector and unknown length. A statistical test such as Hotelling's one-sample T-squared can be used to test the hypothesis that the discrepancies between the observed vector (equal to the mean vector of a sample of vectors) and the theoretical vector may or may not be attributed to chance.

The task of the VF procedure is to estimate the length of the target vector, which, as shown above, corresponds to its contribution to the observed distribution. This can be accomplished by projecting the data vector onto the target vector. This operation is equivalent to rotating the vector space to align one of the axes with the target vector, thus projecting the data vector onto the new axis (see Figure 6). The length of

---

Insert Figure 6 About Here

---

the target vector is equal to the length of the data vector multiplied by the cosine of the angle between the observed and the target vectors. Therefore, the length of the target vector will depend on its orientation. This orientation can be chosen _a priori_, on the basis of knowledge of the spatial distribution of the target component, or of some standard experimental procedure known to elicit the target component. An alternative procedure is to select the orientation of the signal vector on the basis of some _post hoc_ statistical procedure (e.g., discriminant analysis, principal components analysis, etc.).

distributions can also be tested.

Given the usual rules of vector arithmetic, an observed data vector can be viewed as the sum of two or more component vectors, one of which can be considered as an error vector. Each component vector will be characterized by its own orientation in the space (corresponding, as shown above, to a specific spatial distribution) and its own length (which is a measure of the weight of each component vector in determining the data vector). The "contribution" of each component to the data vector is equal to the length of the component vector. Several different procedures to estimate either the length of the component vectors, their orientation, or both, are available (Gratton et al, Note 2). These procedures can be labelled as Vector Decomposition. In the simplest case (the Vector Filter), a single component, with known spatial distribution, is considered responsible for the spatial distribution observed at a given time point and discrepancies between this expected distribution and the observed distributions are attributed to sampling error or noise. A brief description of this procedure is given in the next section.

3.4.4.1 An Application: The Vector Filter

The purpose of the Vector Filter technique (VF, Gratton et al, Note 2) is to determine the amount of the activity recorded with a multiple electrode montage at a given time point that can be attributed to a particular target component, defined *a priori* by the investigator. The target component is defined in terms of a spatial distribution which can be represented as a vector in a multidimensional space. All the activity that cannot be attributed to the target component is defined as error. Orthogonality between the orientations of the target vector and of the error

data vector is a measure of the total activity recorded at all the electrode locations, independent of their relative value or sign. The orientation of the data vector relative to the dimension axes (recording sites) is determined by the relative amplitude and polarity at the different electrode sites, independent of the total activity recorded. Therefore, when a polar notation is adopted to describe the data vector, the information concerning the spatial distribution at each time point can be isolated and expressed by a series of angles between the vector and arbitrary reference axes. Figure 5 shows an example of this notation.

Insert Figure 5 About Here

This approach yields two important benefits: information about the spatial distribution at any given time point can be quantified, and any imaginable spatial distribution can be represented by an orientation in the vector space. In other words, a combination of angles (with the reference axes) in the vector space defines a given spatial distribution.

It is therefore possible to measure the degree to which the spatial distribution observed at a given time point compares with a distribution defined *a priori*. This relationship is described by the cosine of the angle between the observed data vector and the vector representing the hypothesized distribution. A first application of VA to the analysis of spatial distribution consists of establishing the vectors of interest in the vector space, computing the angle with the observed data vectors, and testing the differences. In such an analysis, the same time point from different trials could enter as a replication factor into an one-sample significance test. The difference between two or more observed spatial

The remainder of this section will be concerned with a description of Vector Analysis, a multivariate approach to representing the spatial distribution of a psychophysiological variable. The procedure is both powerful (in that all the available information is used) and efficient (in that signal and noise are clearly distinguished).

3.4.4 Multivariate Approach to Spatial Analysis

Vector Analysis (VA) is a multivariate procedure proposed by Gratton, Coles and Donchin (Note 2) to quantify information about the spatial distribution of a psychophysiological variable. Spatial distribution is here defined as the polarity and relative amount of activity observed at any number of electrode sites, independent of the absolute size of this activity. VA estimates the portion of the activity recorded at several different electrode locations which can be attributed to one or more components, defined in terms of spatial distribution. Therefore, VA defines the signal as one or more components characterized by a specific spatial distribution, and the noise as the remaining variance.

VA treats the voltage values of the electrode locations at a given time point as being the elements of a vector (the data vector). Thus, there is one vector for each time point, and within each vector there is a value for each recording site. This voltage x electrode arrangement contrasts with the usual voltage x time representation. The data vector can be represented geometrically in a space (the vector space) having one dimension for each recording site. Any vector may be characterized by its length and its orientation when plotted in the Euclidean space defined by the dimensions.

VA uses a specific multivariate approach to data reduction such that a univariate approach to inference testing may be employed. The length of the

4.2.4 Time Series Statistics: Methods to Partition Variance

By viewing psychophysiological variables as a time series (i.e., voltage x time functions) and by viewing experimental procedures as a method of partitioning variance we may arrive at two insights into the construct of "variance." First, the variance associated with the "treatment" must be partitioned from the variance associated with the background physiological activity. This procedure is necessary since physiological activity is omnipresent and physiological responses must be evaluated against a varying, rather than constant baseline. Second, the variance of any physiological process is not uniquely determined by any one specific physiological mechanism. Virtually all physiological response systems represent the result of antagonistic mediators which reflect the organism's quest to maintain dynamic homeostasis. Therefore, the variance of the physiological process contains "component" variances representing potentially independent mechanisms. Thus, time series methods may be useful in partitioning the variance of the complex physiological response patterns into components. Moreover, it is possible that the statistical behavior of the components will be different; that is, different components will be differentially sensitive to various manipulations.

The above discussion leads to a revised conceptualization of the physiological response pattern in the psychophysiological experiment. Most physiological response patterns may be conceptualized as the sum of two uncorrelated processes: a baseline trend and an ensemble of rhythmic influences which are superimposed on the baseline trend. The impact of a stimulus or psychological state may reliably influence either or both "component" physiological processes. To complicate matters, the constituent rhythmic components may be manifestations of different underlying

neurophysiological processes. For example, in heart rate there are two obvious rhythms: one modulated at the respiratory frequency (i.e., respiratory sinus arrhythmia); the second, an oscillation at a slower frequency, appears to represent the influence of the rhythmic oscillation of blood and cerebral spinal fluid, since the same rhythm is observed in vasomotor activity, blood pressure, and cerebral spinal fluid.

Time domain approaches focus on evaluating changes in trend as an indicator of the impact of the stimulus manipulation. These methods tend to remove the background rhythmic activity by averaging across trials. The averaging method assumes that the phase relationship between the underlying rhythmic background activity and the stimulus is identically and independently distributed. Thus, when the data are averaged across trials the rhythmic background activity will average to zero. This assumption, of course, is only tenable in experiments in which the timing of stimulus presentation is independent of the physiological process. In self-paced experiments, it is highly unlikely that a rhythmic component of the background physiological activity is not phase related to the self-initiated trial onset, since behavior is neurophysiologically mediated.

In contrast, frequency domain approaches tend to focus on describing the rhythmic components of the background physiological activity which are superimposed on the trend. Thus, it appears that the frequency domain approach tends to evaluate the component of variance which is treated as "error" variance in the time domain approach. Moreover, appropriate implementation of many frequency domain techniques requires that the trend be removed prior to partitioning of the variance into frequency specific components. Frequency domain approaches tend to be associated with spectral analysis technology. The theories underlying the spectral technology have

been, for the most part, developed for "stationary" data sets (Chatfield, 1975). (A time series is said to be stationary when the mean value and autocovariance function are independent of time.) Application of the spectral technology to nonstationary data will result in potentially unreliable and uninterpretable spectral density estimates.

Although any data set which is described in the frequency domain may be represented in the time domain or vice versa, the two approaches do not provide identical information. For example, in the above discussion, we described the primary emphases of time domain (i.e., the description of trend) and frequency domain (i.e., the description of rhythmic activity) approaches. In both approaches the data set typically is modified prior to analysis. In the time domain approach, the data have been "smoothed" to remove the variance associated with background activity. In the frequency domain approach, the data have been "detrended" to provide a stationary data set with a constant baseline. However, any time series, which in our examples would be voltage X time functions, could be described via frequency domain spectral technology in terms of the sum of spectral density estimates and could be "reconstituted" into the original time series with knowledge of the spectral density estimates and the phase relationships among the constituent frequency components. The time domain autocorrelation approach and the frequency domain spectral approach are merely transformations of each other, although time domain models are more likely to be used to describe changes in trend and frequency domain models to describe changes in the constituent frequency components.

The above discussion is relevant since most physiological systems monitored by psychophysiologists tend to have both aperiodic (i.e., trend) and periodic (rhythmic) components. For example, with heart rate we have the basic problem of the directional heart rate responses associated with motor and cognitive function being superimposed on the naturally occurring respiratory sinus arrhythmia. In the case of heart rate, most psychophysiological investigations attempt to maximize the impact of the stimulus or psychological state on the trend. This is done by averaging and treating the rhythmic oscillations as background "error". Similarly, averaging across trials minimizes the background oscillations in electrodermal activity. However, in the case of the EEG recordings it is the periodic characteristics which are emphasized and it is the trend which is filtered from the data set and treated as "error". In both situations, the assumption is made that the physiological response "component" (i.e., trend or periodic) is a sensitive index of the psychological process being monitored. However, it is conceivable that there may be situations in which the "level" of the output of the physiological system manifested in a change in "trend" may be unresponsive to the manipulation, while the treatment effect may be easily observed in a change in the pattern--or vice versa.

4.3 Constraints and Limitations of Sampling Procedures

4.3.1 Physiological Activity: Continuous Processes

Sensitive evaluations of physiological activity must necessarily include sophisticated techniques to evaluate pattern and change. The quantification strategy that the researcher employs in psychophysiological research must rely on an _a priori_ definition of the response parameters being investigated. In most psychological research,

background spontaneous activity is considered unimportant. Meaningful responses can be easily identified as a discrete change in the ongoing activity of the system. However, when investigating physiological processes, it is clear that most physiological systems function continuously. Although we can easily identify the occurrence of many discrete behavioral responses, meaningful physiological responses are often much more difficult to define and isolate. One must assume that virtually every physiological system is continuous even though the measurable datum is manifested at discrete times (e.g., heart beats).

4.3.2 Physiological Activity: Discrete Processes

Although the underlying physiological processes are assumed to be continuous, the prevalent quantification strategies necessitate estimates of the physiological activity at discrete points in time. There are two reasons for this procedure: one, most analytic methods are based upon statistical models in which the continuous process is sampled at sequential points in time; and two, the prevalent quantification techniques associated with digital computers necessitate time-dependent sampling. Thus, although many physiological processes are continuous, the statistical and computer technologies generally force the researcher into quantifying and analyzing the voltage x time functions as discrete processes sampled at sequential points in time.

How fast should one sample continuous processes? The sampling rate or "time window" must be fast enough to accurately describe the variance of the process. The decision regarding sampling rate requires an a priori understanding of the physiological response system being monitored. If relevant information is encoded in a periodic component of the physiological

process with a duration shorter than twice the sampling interval, then the sampled data set will not convey the relevant information. For example, if peripheral vasomotor activity is being sampled from a finger at a rate slower than the heart rate, the variance in vasomotor activity associated with the beating of the heart will be "aliased" or "folded back" on a slower periodicity. The fastest frequency about which we can derive meaningful information from a data set is called the "Nyquist" frequency. The Nyquist frequency is one-half the sampling frequency.

To illustrate the impact of sampling too slowly, consider sampling a 60 Hz pure sine wave 30 times per second. Because the signal would always be at the same point in the cycle when sampled, the samples would all have the same value, implying that no signal is present. Sampling this signal 60 times per second would still yield a flat line. Sampling slightly faster than that would mean measuring successive portions of the cycle, implying a very slowly changing sine wave. For example, sampling a 60 Hz signal 70 times per second would yield a time series of discrete values resembling a pure 10 Hz signal. Indeed, the investigator could not distinguish true 10 Hz activity from "aliased" 60 Hz activity. Only if the 60 Hz signal were sampled 120 or more times per second would the 60 Hz signal not be distorted.

As a more complex example, imagine that there are three physiological variables (i.e., heart rate, respiration rate, and finger vasomotor activity) which are being sampled at a rate of once per second. In this example the heart is beating at 90 beats per minute (i.e., 1.5 beats per second or 1.5 Hz), and the breathing frequency is 15 times a minute (i.e., one breath every 4 seconds or .25 Hz). Sampling each variable once per second, the fastest periodic process we can evaluate in each of the

variables is a process that is slower than one oscillation every 2 seconds or .5 Hz. This does not cause any serious problems with the respiration series since the breathing is slower than the Nyquist frequency of .5 Hz. Similarly in the cardiac system, the fastest periodic activity is the respiratory sinus arrhythmia at the frequency of breathing. However, although vasomotor activity exhibits rhythmic processes at the respiratory frequency and at even slower frequencies, it also oscillates at the frequency of the heart beat since the flow of blood to the periphery is changing on each systole and diastole. Therefore, the peripheral vasomotor activity should exhibit a rhythm of approximately 1.5 Hz (i.e., 90 per minute) concordant with the average heart rate. However, if the vasomotor activity is sampled only once per second, what happens to the variance associated with this fast oscillation? The variance associated with the fast oscillation will be "folded back" and added to the variance of frequencies slower than the Nyquist frequency (which in this case is one-half the 1 Hz sampling rate or .5 Hz). These lower frequencies are said to be "aliased". The same problem will exist if these variables are sampled every 500 msec when the average heart rate is about 90 beat per minute. In this example, the frequency decomposition (spectrum) of the vasomotor time series will result in a periodic component at a frequency slower than breathing, a second "peak" at the breatning frequency, and a third "peak" at a frequency faster than breathing. This faster frequency does not represent a true neurophysiological process, but rather the impact of an inappropriate sampling rate. In this example it would be necessary to sample at 3 Hz or faster (at least twice the 1.5 Hz heart rate) to prevent aliasing. To decompose the rapidly changing vasomotor waveform into frequency components accurately, or to be sensitive to short latency changes in amplitude, it

would be preferrable to sample more than three times a second.

The dangers of inappropriate sampling rates are clear, but how does one avoid these problems? If one were interested in the relationship among various physiological variables, such as heart rate and respiration, it would be necessary to sample the activity of all variables at least twice the frequency of the fastest variable. Note that the problem of aliasing is not problematic solely in the frequency domain--one can see the inappropriate interpretations or loss of relevant information in the time domain, if slow sampling results in not detecting the response component which is sensitive to the stimulus. Fundamentally, sampling a "continuous" process necessitates an understanding of the periodic components and response latencies of the physiological system being studied.

4.3.3 Physiological Activity: Point Processes

Some physiological processes are, by their nature, events which may be characterized as binary--categorized as "occurring" or "not occurring." These processes are called point processes. For example, the beating of the heart may be operationalized as a "binary" event indicated by the occurrence of the R-wave. Similarly, single-unit activity in the central nervous system is characterized by "spikes" and "inter-spike intervals." Point processes pose special statistical problems. The primary problem arises when attempting to sample a point process at equal intervals in time (e.g., second by second). Time series texts (e.g., Gottman, 1981) deal primarily with equal time sampling of continuous processes. Fortunately, this is not problematic with many physiological processes since they may be represented as continuous voltage x time functions. However, how does one deal with processes such as heart period and the ensemble of

processes temporally determined by the beating of the heart? Although blood pressure changes are time locked to the beating of the heart, is it legitimate to view blood pressure as a continuous process and sample at equal time intervals? Moreover, how would one estimate the duration of any specific cardiac cycle component (e.g., P-R interval) across time? These questions have never been adequately discussed in the psychophysiological literature and can be reduced to two points: one, how does one sample "event" related physiological data in equal time intervals; two, how frequently must one sample event-related physiological data?

Although Bartlett (1963) provides a method for performing spectral analysis on the "interval" characteristic of binary data, it is of little use to the psychophysiologist. The reasons are self-evident, since the data are assumed stationary for this analysis. Recall the above arguments that the spectral analysis of "nonstationary" time series provides uninterpretable estimates of the spectral densities and that physiological processes tend to be "nonstationary" time series. It is, therefore, necessary to "detrend" the interval time series to generate a data set which is at least "weakly" stationary. (A process is called weakly stationary if its mean is constant and its autocovariance function depends only on lag: Chatfield, 1975.) Moreover, even in the time domain, equal time interval estimates are necessary for assessing trends. Since most methods of detrending data to produce "stationary" time series for frequency domain analyses are actually time domain methods which have been developed for equal time sampling of continuous process, it is necessary to generate an estimate of the point process at equal points in time.

There are a variety of methods that may be used to generate an estimate of a point process at equal points in time, such as interpolation, weighting, and sampling. Each method has its own unique characteristics. An important requirement is to make the "time window" short enough to map into the temporal variability of the process. If the time window is longer than twice the shortest inter-event interval, then the time window may smooth or alias a component of the variance of the process. In the case of heart period, it is necessary to estimate the heart period in sequential intervals of approximately one-half the duration of the fastest heart period. By estimating the heart period process at sequential intervals which are shorter than half the duration of the fastest heart period, the variance of the heart period process will be preserved in the transformed data set. Moreover, the transformed data set will now be amenable to time-domain detrending and filtering techniques as well as spectral analysis techniques.

4.4 Conclusion

The investigator should consider the relative assumptions, advantages, and disadvantages of time-domain and frequency-domain techniques. Attention to periodicities in a time series, rather than to trends alone can enhance our understanding of psychophysiological processes.

5. Inference Testing

5.1 Introduction

Inference testing involves procedures which evaluate the probable validity of statements about one set of phenomena, where those statements are based on knowledge about a second set of phenomena. The inference being tested may be inductive (one knows real-world event X which appears to be generalizable to principle Y), deductive (one entertains theory Y which predicts real-world event X), or some elaborate combination of these.

Both inductive and deductive inferences typically contribute to a psychophysiological experiment. First, a general concern or hypothesis is stated, and a highly specific instance of it is studied (deduction). Straightforward algebraic manipulations might then be performed on the resulting voltage x time functions to evaluate whether the claim of the hypothesis was manifested in the data obtained. These manipulations produce "descriptive statistics", merely summarizing the data in some highly specified way. Within the confines of a particular experiment, the validity of a hypothesis is tested by inspection of the data--inferential statistical tests are unnecessary. Sections 3 and 4 of this chapter catalog such algebraic procedures, ranging from the computation of a sample mean to PCA.

The investigator is rarely content merely to evaluate the validity of the hypothesis in the specific case alone, however, because the purpose is to confirm the original generalization or to derive new generalizations from initial ones. Thus, the experimenter is likely to attempt to apply one set of concepts to a real-world procedure (deduction), the results of which can then be used to infer new concepts (induction). "Inferential statistics" are those used in this way to evaluate the generalizability of the findings of a particular experiment. Such statistics address the extent to which

findings in a specific case can be expected to hold for some superset of similar cases, versus the alternative that specific results were merely the result of variations in the phenomena not accounted for by the theory under consideration.

As noted earlier in this chapter, the algebraic manipulations which raw data often endure are not easily localized in a single stage of analysis. Just as it was artificial to distinguish between signal extraction (Section 3.2) and data reduction (Section 3.3), so it is artificial to segregate data analysis (Sections 3 and 4) and inference testing. However, while particular techniques straddle such boundaries, the logical distinction between description and inference is essential. The investigator's statistical options become severely curtailed when moving from descriptive/exploratory to inferential analysis.

There are numerous texts on the general use of inferential statistics (e.g., Hays, 1981; Myers, 1979; Winer, 1971). Rather than a complete user's guide to statistical inference, the remainder of this section will provide a sampling of issues of particular relevance to the psychophysiologist, highlighting assumptions of statistical tests, common violations of those assumptions, and remedial solutions.

5.2 Univariate Analysis of Variance (ANOVA)

Two issues arise when traditional ANOVA as an inferential process is applied to psychophysiological data. One issue concerns the need to study phenomena independently of pre-existing basal levels. The use of analysis of covariance (ANCOVA) and of change scores has been particularly controversial. The other issue, the assumption of homogeneity of covariance, derives from the special constraints on ANOVA when repeated

measures are used, either alone or crossed with between-subjects variables (mixed-model ANOVA).

Both issues arise because the psychophysiologist typically studies the time course of voltage x time functions in the context of changing inputs from independent variables. If one wishes to measure acute, event-related responses, variations in average or basal level may contribute a major source of statistical noise (variance not controlled by factors in the ANOVA table appearing in the error term). Alternatively, such variance may constitute the main phenomenon of interest, if one studies slower, homeostatic actions.

Clearly, it is valuable (though not always possible) for the investigator to determine, *a priori*, what are the likely sources of variance in the dependent measure. Gaining experimental control over variance is inherently preferable to attempting to assert *post hoc* statistical control.

5.2.1 Analysis of Covariance (ANCOVA)

Analysis of covariance is commonly the method of choice for *post hoc* removal of undesired sources of variance. Unfortunately, it is often difficult to achieve such statistical control over undesired sources of variance without systematically distorting the data of interest. For example, it has been argued that ANCOVA is not valid in the very situation for which it is intuitively most appealing (see Chapman & Chapman, 1973, pp. 82-83; Lord, 1967). These authors claim that ANCOVA is legitimate only if two requirements are met: the regression slopes of the dependent variable on the covariate must be the same for each level of the independent variable, and the mean value of the covariate must be the same for each level of the independent variable. As an illustration, assume an experiment

in which heart rate (HR) during imagery is believed to vary systematically as a function of imagery ability, a between-subjects factor. To complicate matters, however, some of the subjects are athletes having resting HR levels as much as 40% below those of other subjects. Relative to the hypothesis of the experiment, this source of variance in HR is merely statistical noise. The investigator wishes to employ resting HR as the covariate in an ANCOVA. In order to permit the use of ANCOVA in this case, the investigator would have to show that, for each level of imagery ability in the design, (1) the regression slopes of imagery HR on resting HR are equal and (2) the mean resting HR levels are equal. It is the latter requirement that is most likely to disappoint the investigator, since it is often such hypothetically irrelevant, a priori group differences that tempt the use of ANCOVA. The former requirement, on the other hand, is less constraining--differences in regression slopes amount to an interaction of experimental variables with the covariate, which may be a meaningful, if unanticipated, result (see Section 5.3).

Opinion on the latter requirement is not uniform (see Benjamin, 1967; Cohen & Cohen, 1975; Lubin, 1965; Overall & Woodward, 1977). It has been argued that ANCOVA may be permissible despite group differences on the covariate, depending on the reason for the difference. Overall and Woodward (1977) argued in favor of ANCOVA in the case where experimental treatments do not affect the covariate and subjects are assigned to experimental groups either randomly, or nonrandomly but on the basis of scores on the covariate.

In the case of non-random group assignment, Cohen and Cohen (1975) suggest that the issue be considered in terms of causality. When it is believed (for theoretical--not statistical--reasons) that the covariate is causally dependent on the independent variable, ANCOVA would not be

though at some cost of statistical power. However, non-parametric approaches have not been developed adequately to accommodate the complex experimental designs often used in psychophysiology. One-way and two-way analogs of standard parametric ANOVA have been described and occasionally appear in the literature (Kruskal-Wallis one-way ANOVA by ranks and Friedman two-way ANOVA by ranks, Siegel, 1956). Methods for testing post hoc comparisons following these analyses exist (Levy, 1979; Marascuilo & McSweeney, 1967). Wilson (1956) offered a more general non-parametric ANOVA analog which is computationally more cumbersome and has not generally been used.

Several other non-parametric statistics deserve consideration when the design is appropriate. The well known Spearman rank-order correlation (Rs) has been generalized to the coefficient of concordance statistic (W; Siegel, 1956). Where Rs reflects the agreement between two sets of rankings, W reflects the agreement among multiple sets of rankings. For example, W can reflect the degree of agreement among judges ranking a series of responses. Intuitively, W is the average of the pair-wise Rs values in the data set. Siegel (1956) presents a test of significance for W. This statistic is highly suitable as a summary statistic computed for individual subjects across multiple physiological dependent measures in multiple situations, providing a measure of response stereotypy. Thus, it is potentially useful as a means of classifying subjects. However, it is not so readily applicable to hypothesis testing about relatively homogeneous populations of subjects, the more common goal in experimental design. Some early studies of respone patterning did use W (e.g., Schnore, 1959), but it has received little attention since then.

association between y and set X. If variable y is then replaced by set Y of several variables, it is canonical correlation which measures the association between set X and set Y. Specifically, canonical correlation analysis seeks a linear combination of the variables in set X and a linear combination of the variables in set Y such that a maximum correlation between these two linear combinations is achieved--that set X controls a maximum amount of the variance in set Y. Cohen and Cohen (1975, Chapter 11) and Knapp (1978) demonstrate that canonical correlation subsumes a wide variety of common univariate and multivariate parametric methods of inference testing, including ANOVA, ANCOVA, MRC, MANOVA, MANACOVA (MANOVA with covariance), and discriminant analysis. Given the common practice of quantifying several types of physiological phenomena from multiple recording sites, this technique appears highly appropriate for psychophysiological inference testing. It combines the benefits of MRC and MANOVA (see above) over traditional ANOVA. However, it faces the same interpretative difficulties described for MANOVA.

In general, multivariate statistics have not been widely adopted in psychophysiology, probably because investigators have not felt the need to go beyond what ANOVA will do for them. When the questions asked and the hypotheses tested no longer fit within such strictures, multivariate methods will have to be dealt with. Conversely, once they become routine, they will undoubtedly influence the questions that are asked.

5.5 Nonparametric Tests

Given the frequency with which psychophysiological data violate the assumptions of parametric statistics, non-parametric statistics would seem a highly appropriate alternative. They typically require fewer assumptions,

experiment-wise error rate and to test hypotheses involving several response systems.

Despite these advantages, MANOVA is rarely used in published psychophysiological research. Besides the relative lack of familiarity most investigators have with the technique, two obstacles probably account for this neglect. A practical obstacle is the greater difficulty of computation and statistical interpretation of MANOVA than of ANOVA, including continuing disputes over choice of test statistic (e.g., Olson, 1976, 1979; Stevens, 1979). Standard statistical packages appear to be improving in this regard. However, a conceptual obstacle is the lack of theory to specify the relationship among multiple physiological dependent variables. On what common scale should heart rate and skin conductance be quantified? How readily can one interpret a multivariate F indicating significant systematic variability somewhere among 100 digitized samples from each of eight EEG sites? Thus, even though the basic phenomena of interest are fundamentally multivariate, psychophysiologists have preferred the narrower, univariate ANOVA approach. It is difficult to evaluate how this understandable restriction of vision constrains the hypotheses proposed and the inferences made. The interested reader may consult Cooley and Lohnes (1971), Press (1972), Tatsuoka (1971), VanEgeren (1973), Wilson (1974), Winer (1971), or Woodward and Overall (1975) for basic discussions of MANOVA.

5.4.2 Canonical Correlation Analysis

The traditional Pearson product-moment correlation coefficient, a measure of linear association between variables x and y, may be generalized in two stages. If variable x is replaced by set X consisting of several variables, multiple regression/correlation can evaluate the

In sum, MRC is potentially of great use to the psychophysiologist as a general, conceptually stimulating method of inference testing. The highly readable text by Cohen and Cohen (1975) is recommended to the ANOVA-oriented investigator seeking to employ the more general methods of MRC, particularly when considering ANCOVA (Footnote 2).

5.4 Multivariate Techniques

5.4.1 Multivariate Analysis of Variance (MANOVA)

Although most inferential statistics used in psychophysiology involve multiple variables, multivariate analysis of variance (MANOVA) is a term normally reserved for a technique which is the extension of the typical univariate ANOVA (multiple independent variables but a single dependent variable) to the simultaneous analysis of multiple dependent variables. MANOVA appears to be highly appropriate for inference testing in psychophysiological research, because measurement of multiple dependent measures is routine. MANOVA has been especially advocated as an alternative to univariate ANOVA when repeated measures are involved (Davidson, 1972; Richards, 1980). In the MANOVA approach to such a design, the levels of the repeated-measures variable(s) in the ANOVA become separate dependent variables.

A particular advantage of MANOVA over ANOVA is that while both assume homogeneity of covariance (see Section 5.2.2), studies have shown MANOVA to be very robust to violations of this assumption, especially if cell sizes are equal (Hakstian, Roed, & Lind, 1979; see Richards, 1980). Furthermore, MANOVA is more sensitive than ANOVA to certain types of small but reliable effects (Davidson, 1972). MANOVA is clearly in a better position to control

involve difficult scientific questions with which the investigator should struggle. Typically in ANOVA, all possible interactions are included in the statistical model and the source table. However, in principle the investigator is free to select, on conceptual grounds, which interactions should be included and which ones left in the error term. Use of an incomplete design is particularly appropriate when an interaction term has little theoretical meaning and when its associated degrees of freedom could be put to better use reducing the mean square error. Similarly, whereas ANOVA generally forces the testing of the significance of each factor against an error term which is the residual after all available sources of variance have been removed (i.e., with a minimum of both error sum of squares and error degrees of freedom--a mixed blessing), MRC permits one to use any of several estimates of error, with potentially greater statistical power, consistent with the original Fisherian emphasis on hierarchical (sequential) rather than simultaneous analysis. Again, the choice among these options should be a conscious decision, not delegated to an ANOVA program. As an example, consider an experiment in which subjects' autonomic response to snake exposure is measured. Should the effect of subject gender be tested before or after snake fear questionnaire score has been partialled out of the autonomic measure, and/or partialled out of the gender variable? The converse question can also be raised. ANOVA normally tests each effect after all other effects have been removed from the dependent variable. In other words, each factor is evaluated with all other factors treated as covariates. The investigator may not always find this to be theoretically appropriate.

association of correlational techniques with correlational designs (and their limitations) unnecessarily constrains appreciation of the MRC approach. Fourth, when used for inference testing, MRC is not conveniently adapted for use in repeated measures or mixed-model designs. Fifth, most computer packages do not include in their MRC routine an easy facility for accomplishing an ANOVA-type analysis, especially with nominal independent variables. Finally, MRC normally requires the investigator to make explicit choices about error terms, interactions, and the priority among independent variables. ANOVA programs typically deny the investigator these often difficult choices, with the benefit of convenience but at the cost of flexibility and, sometimes, statistical power.

The last point bears expansion. It must be understood that, both conceptually and algebraically, MRC is a superset of ANOVA. Indeed, such ANOVA complexities as trend analysis, ANCOVA, planned comparisons, continuous independent variables and their interactions, and unequal cell sizes can be handled very routinely within MRC. For example, non-parallel regression slopes which preclude ANCOVA (see Section 5.2.1) amount to an analyzable interaction effect in MRC--that is, a problem becomes a significant finding. Similarly, complex planned comparisons are easily tested using appropriate "dummy coding" of the levels of the independent variable(s).

As one gains experience in the use of MRC to accomplish inference testing of data collected in a traditional ANOVA-type experimental design, one realizes the extent to which most ANOVA packages encourage the investigator to ignore certain basic statistical questions. Specifically, the inclusion or exclusion of particular interaction terms in the statistical model and the pairing of independent variables and error terms

different types of ANOVA designs and emphasizes the simplicity of determining power. A more detailed treatment is available in Cohen (1977).

5.3 Multiple Regression/Correlation (MRC)

Traditionally, the correlational approach in psychology has been associated with psychometric test construction and with studies of individual differences and clinical phenomena. In these cases, independent variables generally vary between subjects and are typically difficult to manipulate experimentally, due to logical, practical, and ethical constraints. Although causality is more difficult to demonstrate with correlational than with experimental designs, there is much to recommend multiple regression/correlation (MRC) as a general data analytic strategy which subsumes ANOVA and ANCOVA (Cohen, 1968; Cohen & Cohen, 1975, especially Section 8.7 and Chapters 9 and 10).

In fact, correlational methods are increasingly used in the processing of psychophysiological data prior to the stage of inference testing. EMCP, Woody filtering, discriminant analysis, PCA, and a number of techniques in the frequency domain employ correlational calculations (see Sections 3.2.2.3 and 4.2).

Several factors have probably contributed to the underutilization of MRC for inference testing in psychophysiology. First, investigators are usually interested in establishing whether a "significant" relationship exists between two variables, rather than in estimating the strength of the relationship, or predicting the exact value of one variable given the other. More conceptually, the specific prediction of a physiological variable on the basis of a psychological variable would be meaningful only given a level of theorizing beyond what is often available. Third, the traditional

paradigm often implicitly guides the decision, with sample size and alpha-level set accordingly.

However, in the opinion of the present authors, statistical power is frequently too low in many psychophysiological studies. While it may be argued that power is low only for weak effects and that perhaps we should normally confine ourselves to seeking strong effects, the issue is rather that investigators too rarely consider the questions of effect size and power explicitly when designing experiments.

As an illustration of low power levels prevalent in psychophysiological research, effect size and power were calculated for a subset of data published in Lang et al (1980). For a simple between-subjects main effect, with 16 subjects in each of two groups--a relatively large sample size for a psychophysiological study--the effect sizes actually obtained for two dependent measures were equivalent to correlations of .43 and .35. These were conceptually important effects and were statistically significant. Assuming on alpha-level of .05 and an effect size of .40 (Cohen and Cohen suggest .3 as "moderate" and .5 as "large" effect sizes when one has no basis for estimate), a total sample size of 32 provides a power level of .64. In other words, Lang et al could expect to find such an effect, upon replication, in only two experiments out of every three attempts. If sample size were reduced to 24, power would fall to .51, or only an even chance of replication.

The point of this illustration is that statistical power is surprisingly low in many psychophysiological studies. Cohen and Cohen (1975) provide an enlightening discussion of power and a straightforward method for its computation. They recommend .80 as a reasonable target value for power in many situations. Koele (1982) discusses the relative power of

(Jennings & Wood, 1976). These factors together undoubtedly explain how rarely the correction factor is used in published research.

The second way to cope with violations of the homogeneity of variance assumption, advocated by Richards (1980), is to do multivariate analysis of variance (MANOVA), rather than the usual repeated-measures, univariate ANOVA. The different levels of the repeated-measures factor in ANOVA become separate dependent variables analyzed simultaneously in MANOVA (see Section 5.4).

5.2.4 Power of the F-test in ANOVA

Statistical power is generally well understood conceptually but is rarely considered quantitatively in the design or evaluation of psychophysiological studies. Power may be defined as the ability to find an effect which actually exists. More formally, if $\underline{B}$ (beta) is the probability of a Type II error (failure to reject a false null hypothesis), then power is $1 - \underline{B}$.

Cohen and Cohen (1975) discuss statistical power as a joint function of three other parameters, such that power is increased when any of the following is increased: sample size, effect size, and $\underline{a}$ (alpha--the probability of a Type I error, rejection of the null hypothesis when it is true). Fixing values of any three of these parameters determines the fourth. Conversely, the appropriate value for any of them, such as sample size, cannot be determined without knowing the other three. Effect size is often most difficult to deal with in experimental design. The magnitude of an experimental effect can intuitively be understood as the value of an equivalent correlation coefficient. Although the investigator's hypotheses may not specify the effect size anticipated, prior experience with a given

evaluated using the most conservative [(1), (N-1)] and the most liberal [(K-1), (K-1) x (N-1)] limits; if the *F* is significant in the first case or non-significant in the latter case, there is no need to calculate the correction factor.

To illustrate the potential impact of the Geisser-Greenhouse correction factor, consider a standard CNV paradigm (Simons, Ohman, & Lang, 1979). The EEG was digitized at 30 Hz, and sets of 15 consecutive points were averaged to one value every half-second. One of the present authors computed the correction factor for this data set to be .19. Thus, using the correction factor would have cost the investigators 80% of their degrees of freedom in analyses using the half-second data points. Using median heart rate data obtained during sequential 30-second periods of an imagery experiment (Lang, Kozak, Miller, Levin, & McLean, 1980), the correction factor was computed to be .35. This higher value (i.e., less heterogeneity) reflects the much longer inter-observation interval in this study than in the CNV example, though about two-thirds of the degrees of freedom would still have been lost had the correction been made.

In sum, the Geisser-Greenhouse correction factor can take a very serious toll on the apparent statistical power of a psychophysiological experiment (though, of course, it merely reclaims the inflated "power" caused by heterogeneity of covariance). In addition, Davidson (1972) has pointed out that with small sample sizes there may be inadequate statistical power in the procedure to detect a violation of the homogeneity assumption. Finally, it has been suggested that the product of the correction factors for the separate main effects be used for testing interactions of repeated-measures factors, but this application has not been fully validated

Violations of the assumption are possible in any repeated-measures design, but almost inevitable in psychophysiological studies analyzing voltage x time functions. Neighboring time points tend to be highly correlated, but samples which are more widely spaced in time will generally be less tightly coupled. The problem is exacerbated when periodicities exists in the signal, such as sinus arrhythmia in heart rate, producing systematic irregularities in the covariance among sample points. Thus, the covariances among pairs of points in a time series will vary as a function of their temporal separation. This problem is clearly larger when "time" is measured in milliseconds between digitized samples than in minutes between trial-blocks or days between sessions. Of course, the critical issue is not the absolute time scale but the stability of the psychophysiological function relative to the sampling interval and the total sample epoch. Neighboring ten-per-second samples of skin conductance level are likely to be much more intercorrelated than ten-per-second samples of EEG.

Two ways of coping with heterogeneity of covariance have been proposed. Jennings and Wood (1976) apprised psychophysiologists of Box's (1954) solution as developed by Geisser and Greenhouse (1958; see also Games, 1975, 1976; Keselman & Rogan, 1980; Keselman, Rogan, Mendoza, & Breen, 1980; McCall & Appelbaum, 1973; Richards, 1980; Wilson, 1974). This method reduces the degrees of freedom used for evaluation of the significance of the $\underline{F}$ statistic. The reduction is proportional to the amount of heterogeneity among the covariances. Assuming K treatment levels and N subjects, the maximum possible reduction is from (K-1) and (K-1) x (N-1) to (1) and (N-1) degrees of freedom. Jennings and Wood (1976) and Myers (1979) provide a formula for calculating the correction factor, known as epsilon or lambda (Footnote 1). These writers point out that the $\underline{F}$ can first be

on response amplitude scores, it was later shown (Benjamin, 1963) that the ALS actually removes LIV effects completely. Essentially, the ALS method is a customized elaboration of ANCOVA. As such, it is vulnerable to the same constraints and debates as ANCOVA (see Section 5.2.1). When the investigator is satisfied that the ALS method is statistically acceptable in a particular data set, it can be useful for standardizing data consisting of multiple dependent measures having different measurement scales, variances, etc. It is particularly appropriate for a close examination of response pattern across situations and measures (e.g., Lacey, 1956; cf. coefficient of concordance, Section 5.5). However, the ALS is rarely used in current research.

The validity of change scores is still arguable (e.g., Benjamin, 1973; Etaugh & Etaugh, 1972; Harris, 1963; Lubin, 1965). The present authors are in sympathy with Benjamin (1973), who argued that one's metric achieves validity because of its theoretical appropriateness, not its statistical purity. Thus, if one's theory actually makes predictions about change scores, one should measure change scores.

5.2.3 The Assumption of Homogeneity of Covariance

Homogeneity of covariance means that the covariance between each pair of repeated-measures factor levels is constant. The assumption of homogeneity of covariance in ANOVA is perhaps less controversial than the requirement for ANCOVA, but it is still frequently violated. For example, Jennings and Wood (1976) reported that 84% of the articles using repeated-measures ANOVA designs in Volume 12 of Psychophysiology (1975) appeared to have ignored the assumption. Violations of the assumption generally bias the *F*-test toward a Type I error (Myers, 1979; Winer, 1971).

will pose an interpretative problem. Perhaps the important lesson to be drawn from such debates is that what constitutes a proper use of inferential statistics is partially a function of the experimental purpose and conceptual framework. Just as a given statistical model is developed under certain assumptions, the importance of a violation of those assumptions rests on the use made of the statistic. Furthermore, particular assumptions differ with respect to the consequences of violation. However, the confidence intervals of traditional statistics are not the only means of testing inferences. Repeated sampling from the superset of cases to which one's first experiment belongs--known as replication and cross-validation--is a respectable alternative.

5.2.2 Change Scores and the Law of Initial Values (LIV)

Wilder (1957) first formulated the Law of Initial Values (LIV), stating that response amplitude is a function of prestimulus level. Assuming an implicit ceiling effect, the prediction is that higher prestimulus levels will be associated with smaller responses. A number of early papers report confirmatory data (e.g., Hord, Johnson, & Lubin, 1964; Lacey, 1956; Sternbach, 1960). The LIV has serious implications for the most popular metric in psychophysiology, the change score. Specifically, the size of the change score may be partially a function of initial level. Clearly, the LIV phenomenon can affect change-score data in ways which are unrelated to the experimental manipulation, generally reducing statistical power. To deal with this problem with autonomic measures, where the LIV problem is most widely acknowledged, Lacey (1956) proposed the Autonomic Lability Score transformation (ALS). While the ALS was intended to take account of the LIV by standardizing the effect of the homeostatic influence

legitimate. Specifically, if covariate C shares variance with independent variable X because X affects C, then removing their shared variance from X unfairly robs X of variance with which X may actually affect dependent variable Y. Thus, ANCOVA in this case would distort the experimental effect of X on Y. On the other hand, if C is causally prior to X, then any shared variance with which they jointly affect Y properly belongs to C, not X. Y would be affected by that source of variance whether or not X were present, so X should not be credited with that variance. In the above example of imagery and heart rate, where group assignment was nonrandom (based on imagery ability), it could be argued that there is no theoretical basis for basal HR determining imagery ability. Thus, the use of ANCOVA in the face of group differences on the covariate could be defended.

In the case of random assignment to groups, Cohen and Cohen (1975) are quite comfortable with ANCOVA. They reason that, although two random samples may by chance differ on the covariate, what matters is the population they represent, about which inferences will be drawn. Randomization assures that the expected (population) differences between samples will be zero, regardless of actual sample differences. Thus, C and X share no variance in the population, even though they may do so by chance in the samples selected. Although the groups differ on C, they will tend to regress toward their population mean (i.e., no difference on C) when they are observed in order to measure Y. Of course, this tendency will be realized only for large samples.

We will not attempt to advocate one or the other of these positions on the validity of ANCOVA in the face of group differences on the covariate. Clearly, the investigator should evaluate whether the two assumptions are met in a given data set and consider whether violation of the assumptions

Dependent variables in psychophysiological studies are usually quantified as continuous variables. However, when the dependent variable in a repeated-measures design is dichotomous (for example, responses might be scored as present or absent, as is sometimes done for skin conductance responses), the usual parametric ANOVA is not in general appropriate (Marascuilo & McSweeney, 1977; Winer, 1971). The non-parametric $\underline{Q}$ statistic (Cochran, 1950) is recommended instead, although under some conditions (including a sufficiently large number of observations), the $\underline{F}$ of ANOVA approximates $\underline{Q}$ (D'Agostino, 1971; Lunny, 1970). $\underline{Q}$ is easily computed and follows a chi-square distribution. $\underline{Q}$ is vulnerable to heterogeneity of covariance in a manner analogous to the $\underline{F}$ statistic, and the Box/Geisser-Greenhouse correction factor for degrees of freedom in the $\underline{F}$-test is appropriate for $\underline{Q}$ (Bhapkar & Somes, 1977; Myers, DiCecco, White, & Borden, 1982). Methods for testing post hoc comparisons following the $\underline{Q}$ analysis exist (Levy, 1979; Marascuilo & McSweeney, 1977). Although developed for a simple subjects x conditions design, $\underline{Q}$ has been extended to certain cases of interaction (Marascuilo & Serlin, 1977). However, $\underline{Q}$ has not been extended adequately to cover the complex designs typical of much of psychophysiology.

5.6 A Few More Caveats

The use of change scores pervades data analysis in psychophysiology. Nevertheless, change scores are notorious for having low reliability. Furthermore, if treatment means and true-score variance are held constant, statistical power is directly related to the reliability of the dependent variable (i.e., inversely related to error-score variance). Thus, change scores seem a poor candidate for inference testing. However, Nicewander and

Price (1978) have recently demonstrated that high reliability is not necessarily optimal for inference testing, largely because true-score variance is often not held constant in experiments. They showed, in fact, that under certain conditions statistical power is paradoxically maximized when dependent measure reliability is minimized. They concluded that the optimal level of reliability depends on the nature of the hypothesis being tested. One caveat would be that the investigator should consider carefully whether, on conceptual grounds, change scores are appropriate in a given study.

A related issue is the direction of statistical inference and the relative reliability of different measurements. Though more often discussed in the clinical realm, this issue arises in psychophysiological research as well. Chapman and Chapman (1973, p. 67) provide a clear example: "If schizophrenics are as inferior to normal subjects on one ability as on another, but the test that is used to measure one of the abilities is more reliable than the test for the other, a greater deficit will be found on the more reliable measure." In psychophysiology, a number of examples could be given. If two conditions produce equal real changes from a third condition, the measured change will be larger for the condition measured with greater reliability. Similarly, if one quantifies heart rate with greater reliability than finger pulse volume, genuinely equal changes in both will yield data indicating a larger change in heart rate than in finger pulse volume. As a final example, if P300 in the event-related brain potential can be measured more reliably (in a statistical sense) at Pz than at Fz, it will be easier to find same-size effects at Pz than at Fz. A second caveat, then, would be that investigators should evaluate the statistical reliability of psychophysiological measures and, in making statistical

inferences, should beware not to confound differences in genuine effects with differences in reliability of measurement.

6. Conclusion

In this chapter, we have reviewed techniques that can be used to analyze psychophysiological measures. We have argued that, in the end, procedures used to measure all psychophysiological functions result in a voltage x time function. For this reason, all analytic techniques can, at least in principle, be applied to any psychophysiological function. Selection of which technique to use must be guided, in part, by the particular question the investigator seeks to answer and, in part, by the nature of the underlying physiological system. We have attempted to indicate the advantages and disadvantages of each technique. By doing this, we hope that investigators will look beyond those techniques that have traditionally been associated with a particular function. We also hope that, in spite of space limitations, we have given enough guidance to enable the interested researcher to make an intelligent selection of a technique. The references we have provided should ensure that users go beyond a "cookbook" approach.

We should emphasize that analytic techniques are, in some sense, only as good as the data to which they are applied. There is clearly no substitute for careful recording procedures and appropriate experimental design.

7. Reference Notes

Note 1. Tucker, L. R. _Note on number of attributes in a battery to support a given number of common factors_. Unpublished manuscript, 1973.

Note 2. Gratton, G., Coles, M. G. H., & Donchin, E. A vector analysis of ERPs, in preparation.

8. References

Ackroyd, M. Digital filtering. London: Butterworth, 1973.

Aunon, J.I., McGillem, C.D., & O'Donnell, R.D. Comparison of linear and quadratic classification of event-related potentials on the basis of their exogenous or endogenous components. Psychophysiology, 1982, 19, 531-537.

Bartlett, M. S. The spectral analysis of point processes. Journal of the Royal Statistical Society, Series B, 1963, 25, 264-280.

Benjamin, L. S. Statistical treatment of the law of initial values (LIV) in autonomic research: A review and recommendation. Psychosomatic Medicine, 1963, 25, 556-566.

Benjamin, L. S. Facts and artifacts in using analysis of covariance to "undo" the law of initial values. Psychophysiology, 1967, 4, 187-206.

Benjamin, L. S. Remarks on behalf of change scores and associated correlational statistics: A response to the Etaughs. Developmental Psychology, 1973, 8, 180-183.

Bhapkar, V. P., & Somes, G. W. Distribution of Q when testing equality of matched proportions. Journal of the American Statistical Association, 1977, 72, 658-661.

Bohrer, R. E., & Porges, S. W. The application of time-series statistics to psychological research: An introduction. In G. Keren (Ed.), Statistical and Methodological Issues in Psychology and Social Sciences Research. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1982.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and covariance between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-494.

Box, G. E. P., & Jenkins, G. M. Time series analysis, forecasting and control. San Francisco: Holden Day, 1976.

Brillinger, D. R. Time series: Data analysis and theory. New York: Holt, Rinehart & Winston, 1975.

Brown, C. C. (Ed,) Methods in psychophysiology. Baltimore: Williams & Wilkins, 1967.

Callaway, E., & Halliday, R. A. Evoked potential variability: Effects of age, amplitude and methods of measurement. Electroencephalography & Clinical Neurophysiology, 1973, 34, 125-133.

Callaway, E., Halliday, R., & Herning, R. I. A comparison of methods for measuring event-related potentials. Electroencephalography & Clinical Neurophysiology, 1983, 55, 227-232.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Carlton, E. H., & Katz, S. Is Wiener filtering an effective method of improving evoked potential estimation? IEEE Transactions in Biomedical Engineering, 1980, 27, 187-192.

Cattell, R. B. The scree test for the number of factors. Multivariate Behavioral Research, 1966, 1, 245-276.

Chapman, L. J., & Chapman, J. P. Disordered thought in schizophrenia. New York: Appleton-Century-Crofts, 1973.

Chatfield, C. The analysis of time series: theory and practice. London: Chapman & Hall, 1975.

Cochran, W. G. The comparison of percentages in matched samples. Biometrika, 1950, 37, 256-266.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70, 426-443.

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1977.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N J: Lawrence Erlbaum, 1975.

Cooley, W. W., & Lohnes, P. R. Multivariate data analysis. New York: Wiley, 1971.

Cook, E. C. FWTGEN--An interactive FORTRAN II/IV program for calculating weights for a non-recursive digital filter. Psychophysiology, 1981, 18, 489-490.

Corby, J. C., & Kopell, B. S. Differential contribution of blinks and vertical eye movements as artifacts in EEG recording. Psychophysiology, 1972, 9 , 640-644.

Cox, D. R., & Lewis, P. A. W. The statistical analysis of series of events. London: Methuen, 1966.

D'Agostino, R. B. A second look at analyis of variance on dichotomous data. Journal of Educational Measurement, 1971, 8, 327-333.

Davidson, M. L. Univariate versus multivariate tests in repeated-measures experiments. Psychological Bulletin, 1972, 77, 446-452.

Dixon, W. T. BMDP biomedical computer programs. University of California, Los Angeles, 1979.

Donchin, E. A multivariate approach to the analysis of average evoked potentials. IEEE Transactions in Biomedical Engineering, 1966, 13, 131-139.

Donchin, E. Discriminant analysis in average evoked response studies: The study of single trial data. Electroencephalography & Clinical Neurophysiology, 1969, 27, 311-314.

Donchin, E., Use of scalp distribution as a dependent variable in Event-Related Potentials studies: excerpts of preconference correspondence. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research, EPA-600/9-77-043, Washington, D.C.: U.S. Government Printing Office, 1978.

Donchin, E., & Heffley, E. Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research. EPA 600/9-77-043, Washington, D.C.: U.S. Government Printing Office, 1978.

Donchin, E., & Herning, R. I. A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event-related potentials. Electroencephalography & Clinical Neurophysiology, 1975, 38, 51-68.

Donchin, E., Tueting, P., Ritter, W., Kutas, M., & Heffley, E. On the independence of the CNV and the P300 components of the human average evoked potential. Electroencephalography & Clinical Neurophysiology, 1975, 38, 449-461.

Duffy, F. H., Bartels, P. H., & Burchfiel, J. L. Significance probability mapping: an aid in the topographical analysis of brain electrical activity, Electroencephalography and Clinical Neurophysiology, 1981, 51, 455-462.

Ertl, J. P. Detection of evoked potentials by zero crossing analysis. Electroencephalography & Clinical Neurophysiology, 1965, 18, 630-631.

Ertl, J. P., & Schafer, E. W. P. Brain response correlates of psychometric intelligence. Nature, 1969, 223, 421-422.

Etaugh, A. F., & Etaugh, C. F. Overlap: Hypothesis or tautology? Developmental Psychology, 1972, 6, 340-342.

Friedman, D. H. Detection of signal by template matching. Baltimore: The John Hopkins Press, 1968.

Games, P. A. Computer programs for robust analyses in multifactor analysis of variance designs. Educational and Psychological Measurement, 1975, 35, 147-152.

Games, P.A. Programs for robust analyses of ANOVA's with repeated measures. Psychophysiology, 1976, 13, 603.

Geisser, S., & Greenhouse, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics, 1958, 29, 885-891.

Glaser, E. M., & Ruchkin, D. S. Principles of neurobiological signal analysis. New York: Academic Press, 1976.

Gottman, J. M. Time-series analysis. New York: Cambridge University Press, 1981.

Graham, F. K. Constraints on measuring heart rate sequentially through real and cardiac time. Psychophysiology, 1978, 15, 492-495.

Gratton, G., Coles, M. G. H., & Donchin, E. A new method for off-line removal of ocular artifact. Electroencephalography & Clinical Neurophysiology, 1983, 55, 468-484.

Hakstian, A. R., Roed, J. C., & Lind, J. C. Two-sample T2 procedure and the assumption of homogeneous covariance matrices. Psychological Bulletin, 1979, 86, 1255-1263.

Hall, R. A., Rappaport, M., Hopkins, H. K., & Griffin, R. B. Peak identification in visual evoked potentials. Psychophysiology, 1973, 10, 52-60.

Harman, H. H. Modern Factor Analysis, (Revised Edition). Chicago: University of Chicago Press, 1967.

Harris, C. W. (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Hays, W. L. Statistics. New York: Holt, Rinehart & Winston, 1981.

Hord, D. J., Johnson, L. C., & Lubin, A. Differential effect of the law of initial value (LIV) on autonomic variables. Psychophysiology, 1964, 1, 79-87.

Horst, R. L., & Donchin, E. Beyond averaging II: Single trial classification of exogenous event-related potentials using stepwise discriminant analysis. Electroencephalography & Clinical Neurophysiology, 1980, 48, 113-126.

Humphreys, L. G., & Montanelli, R. G. An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research, 1975, 10, 193-206.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 1980, 17, 259-273.

Jennings, J. R., Tahmoush, A. J., & Redmond, D. P. Non-invasive measurement of peripheral vascular activity. In I. Martin & P. H. Venables (Eds.), Techniques in psychophysiology. Chichester, England: Wiley, 1980.

Jennings, J. R., & Wood, C. C. The epsilon-adjustment procedures for repeated-measures analyses of variance. Psychophysiology, 1976, 13, 277-278.

Jennrich, R. I. Stepwise discriminant analysis. In K. Enslin, A. Ralston and H.S. Wief (Eds.), Statistical methods for digital computers. New York: Academic Press, 1977.

John, E. R., Ruchkin, D. S., & Villegas, J. Experimental background: Signal analysis and behavioral correlates of evoked potential configurations in cats. Annual New York Academy of Science, 1964, 112, 362-420.

John, E. R., Ruchkin, D. S., & Vidal, J. J. Measurement of event-related potentials. In E. Callaway, P. Tueting and S.H. Koslow (Eds.), Event-Related Brain Potentials in Man. Academic Press, New York, 1978.

Kaiser, H. F. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 1960, 20, 141-151.

Keselman, H. J., & Rogan, J. C. Repeated measures F tests and psychophysiological research: Controlling the number of false positives. Psychophysiology, 1980, 17, 499-503.

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. Testing the validity conditions of repeated measures F tests. Psychophysiological Bulletin, 1980, 87, 479-481.

Knapp, T. R. Canonical correlation analysis: A general parametric significance-testing system. Psychological Bulletin, 1978, 85, 410-416.

Koele, P. Calculating power in analysis of variance. Psychological Bulletin, 1982, 92, 513-516.

Kubayashi, H., & Yaguchi, K. A statistical method of component identification of average evoked potentials. Electroencephalography & Clinical Neurophysiology, 1981, 51, 213-214.

Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. Annals of the New York Academy of Sciences, 1956, 67, 123-164.

Lang, P. J., Kozak, M. J., Miller, G. A., Levin, D. N., & McLean, A. Emotional imagery: Conceptual structure and pattern of somatovisceral response. Psychophysiology, 1980, 17, 179-190.

Levy, K. J. Nonparametric large-sample pairwise comparisons. Psychological Bulletin, 1979, 86, 371-375.

Lord, F. M. A paradox in the interpretation of group comparisons. Psychological Bulletin, 1967, 68, 304-305.

Lubin, A. Book review of Problems in measuring change, Ed. by C. Harris. American Journal of Psychology, 1965, 78, 324-327.

Lunny, G. H. Using analysis of variance with a dichotomous dependent variable: An empirical study. Journal of Educational Measurement, 1970, 4, 263-269.

Malmstadt, H. V., Enke, C. G., & Crouch, S. R. Electronic measurements for scientists. Mendo Park: W. A. Benjamin, 1974.

Marascuilo, L. A., & McSweeney, M. Nonparametric post hoc comparisons for trend. Psychological Bulletin, 1967, 67, 401-412.

Marascuilo, L. A., & McSweeney, M. Nonparametric and distribution free methods for the social sciences. Monterey, CA: Brooks/Cole, 1977.

Marascuilo, L. A., & Serlin, R. Interactions for dichotomous variables in repeated measures design. Psychological Bulletin, 1977, 84, 1002-1007.

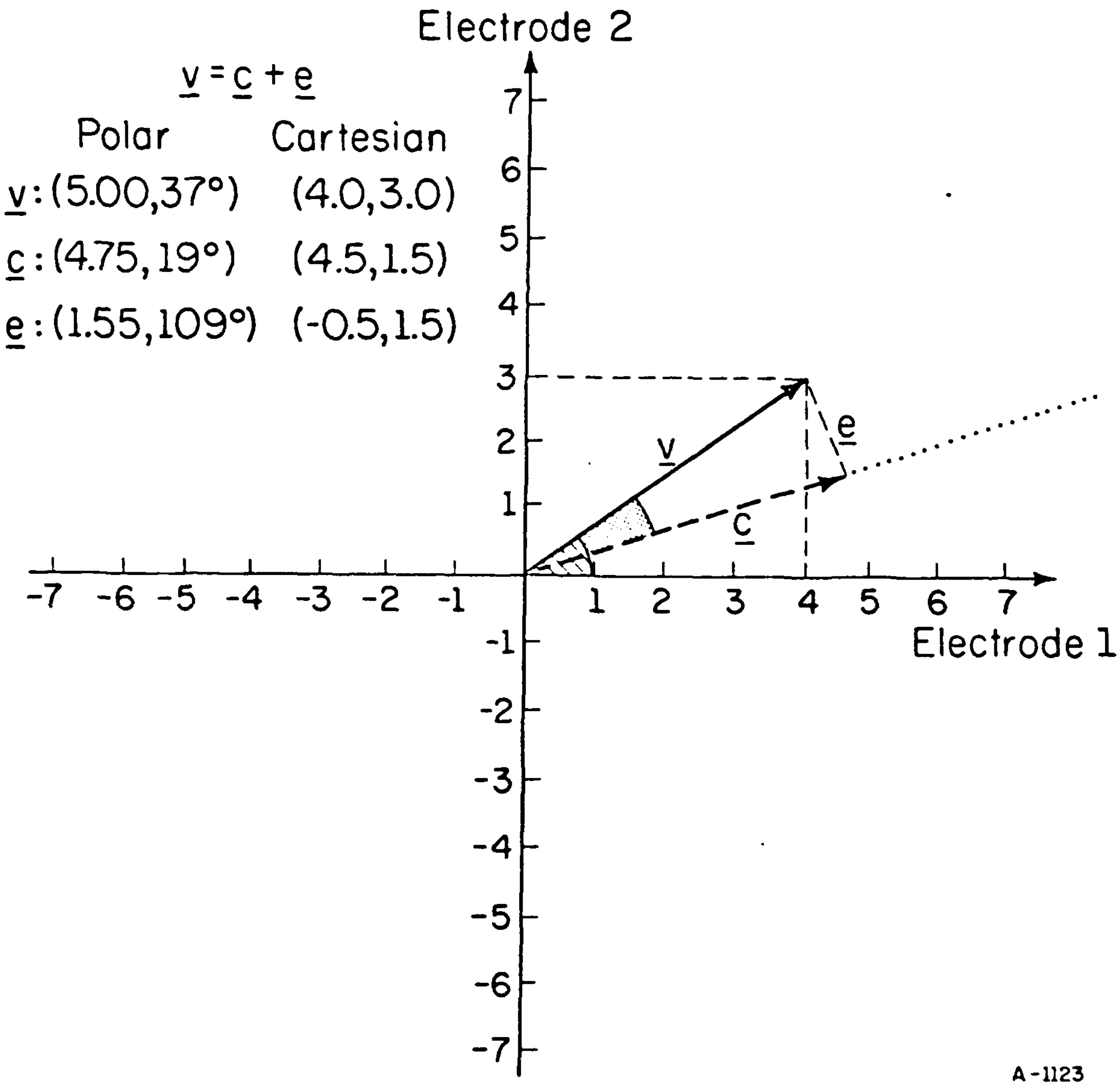
Martin, I., & Venables, P. H. (Eds.) Techniques in psychophysiology. Chichester, England: Wiley, 1980.

McCall, R. B., & Appelbaum, M. I. Bias in the analysis of repeated-measures designs: Some alternative approaches. Child Development, 1973, 44, 401-415.

McGillem, C. D., Aunon, J. I., & Childers, D. G. Signal processing in evoked potential research: Applications of filtering and pattern recognition. Critical Reviews in Bioengineering, 1981, 6, 225-265.

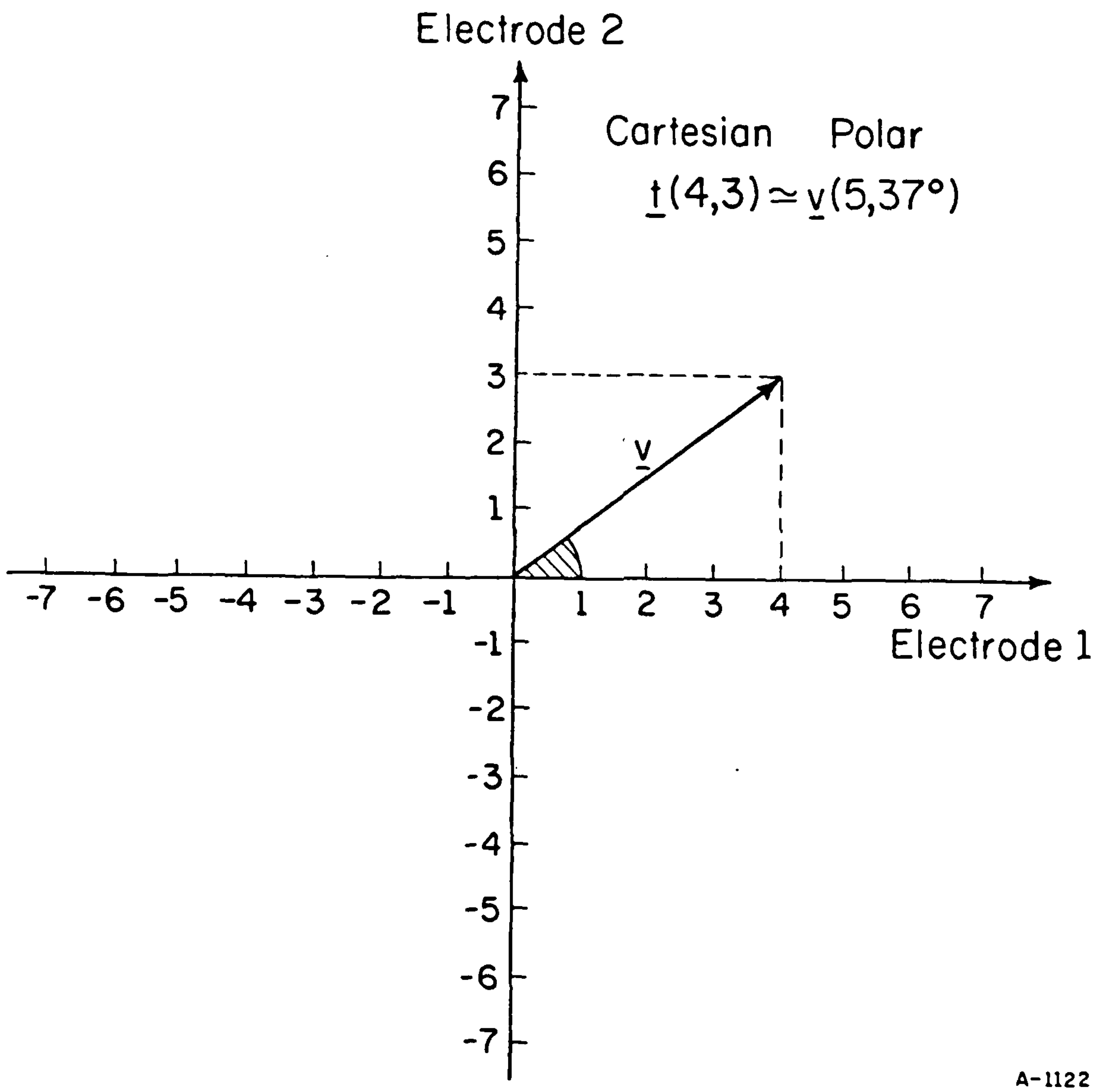
McGillem, C. D., Aunon, J. I., & O'Donnell, R. D. Computer classification of single event-related potentials. Psychophysiology, 1981, 18, 192 (abstract).

Montanelli, R. G., & Humphreys, L. G. Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. Psychometrika, 1976, 41, 341-347.



Electrode 2
v = c + e
Polar
Cartesian
v: (5.00,37°)
(4.0,3.0)
c: (4.75,19°)
(4.5,1.5)
e: (1.55,109°)
(-0.5,1.5)
v
c
e
Electrode 1
A-1123

Figure 6



Electrode 2
Cartesian Polar
$\underline{t}(4,3) \simeq \underline{v}(5,37°)$
$\underline{v}$
7
6
5
4
3
2
1
-1
-2
-3
-4
-5
-6
-7
-7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7
Electrode 1
A-1122

Figure 5

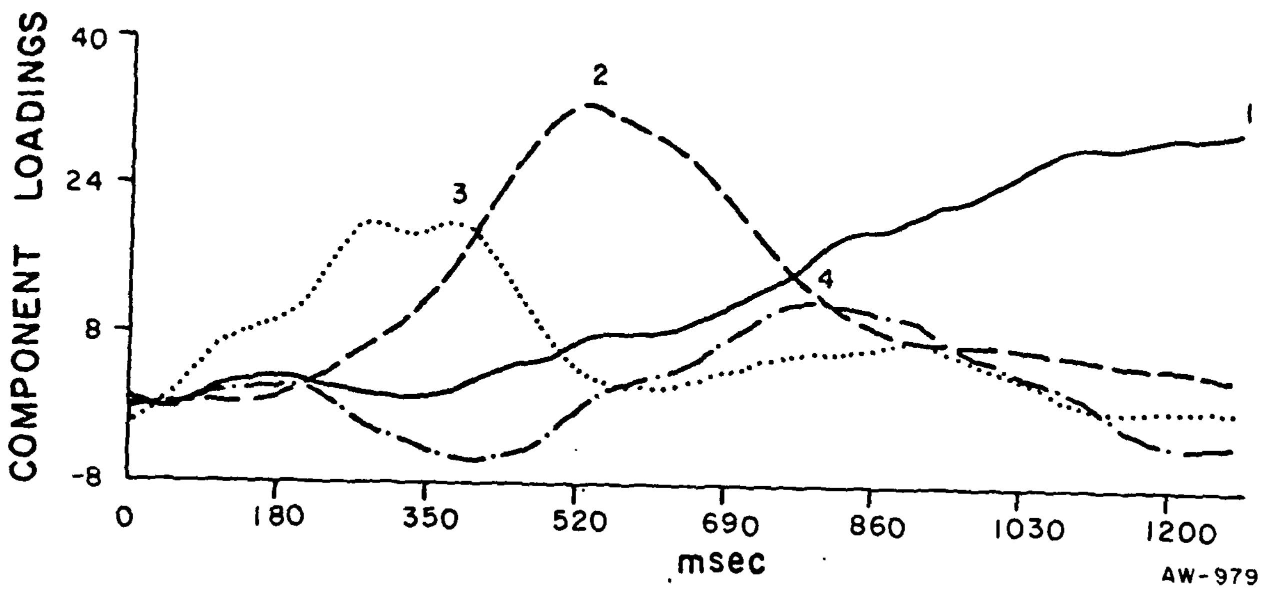
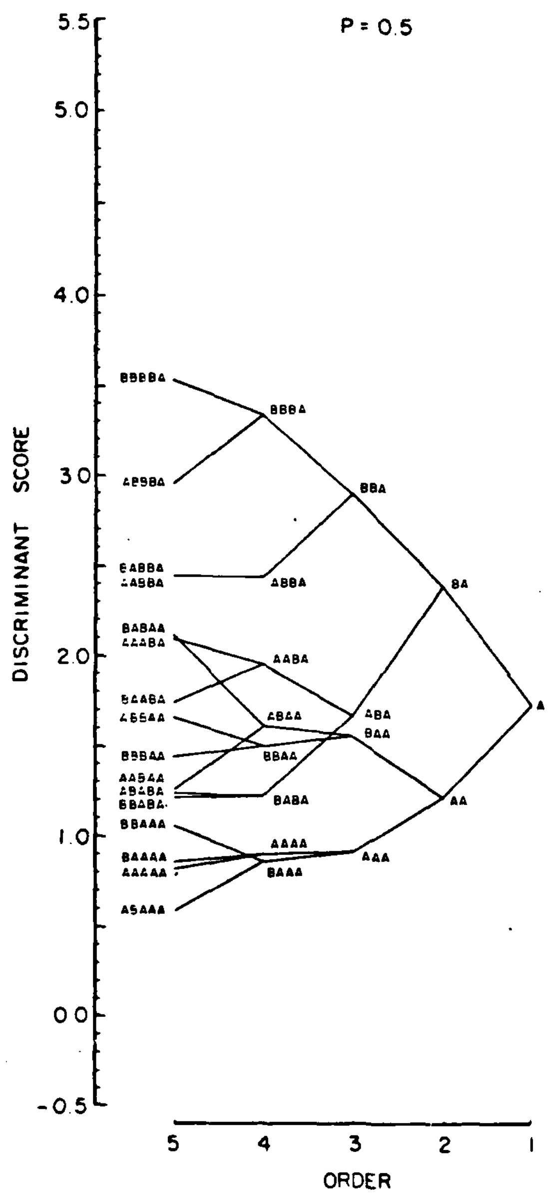

COMPONENT LOADINGS
40
24
8
-8
0
180
350
520
690
860
1030
1200
msec
1
2
3
4
AW-979

Figure 4



P = 0.5
DISCRIMINANT SCORE
ORDER
5.5
5.0
4.0
3.0
2.0
1.0
0.0
-0.5
5
4
3
2
1
BBBBA
BBBA
BBA
BA
ABBA
AABA
ABAA
ABA
BAA
BBAA
BABA
AA
AAAA
AAA
BAAA

Figure 3

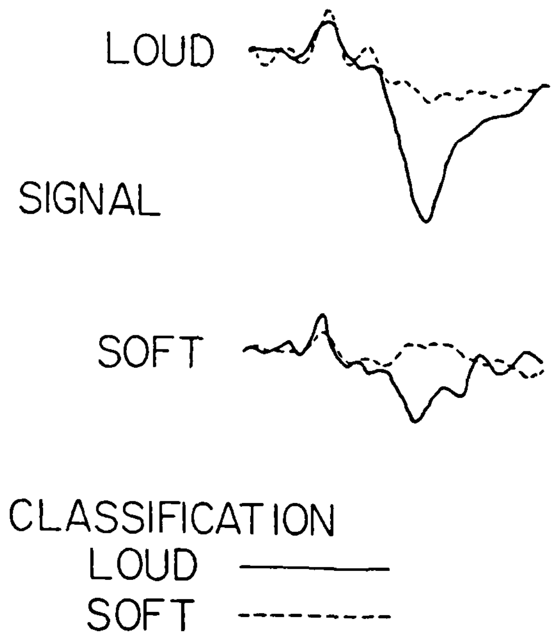
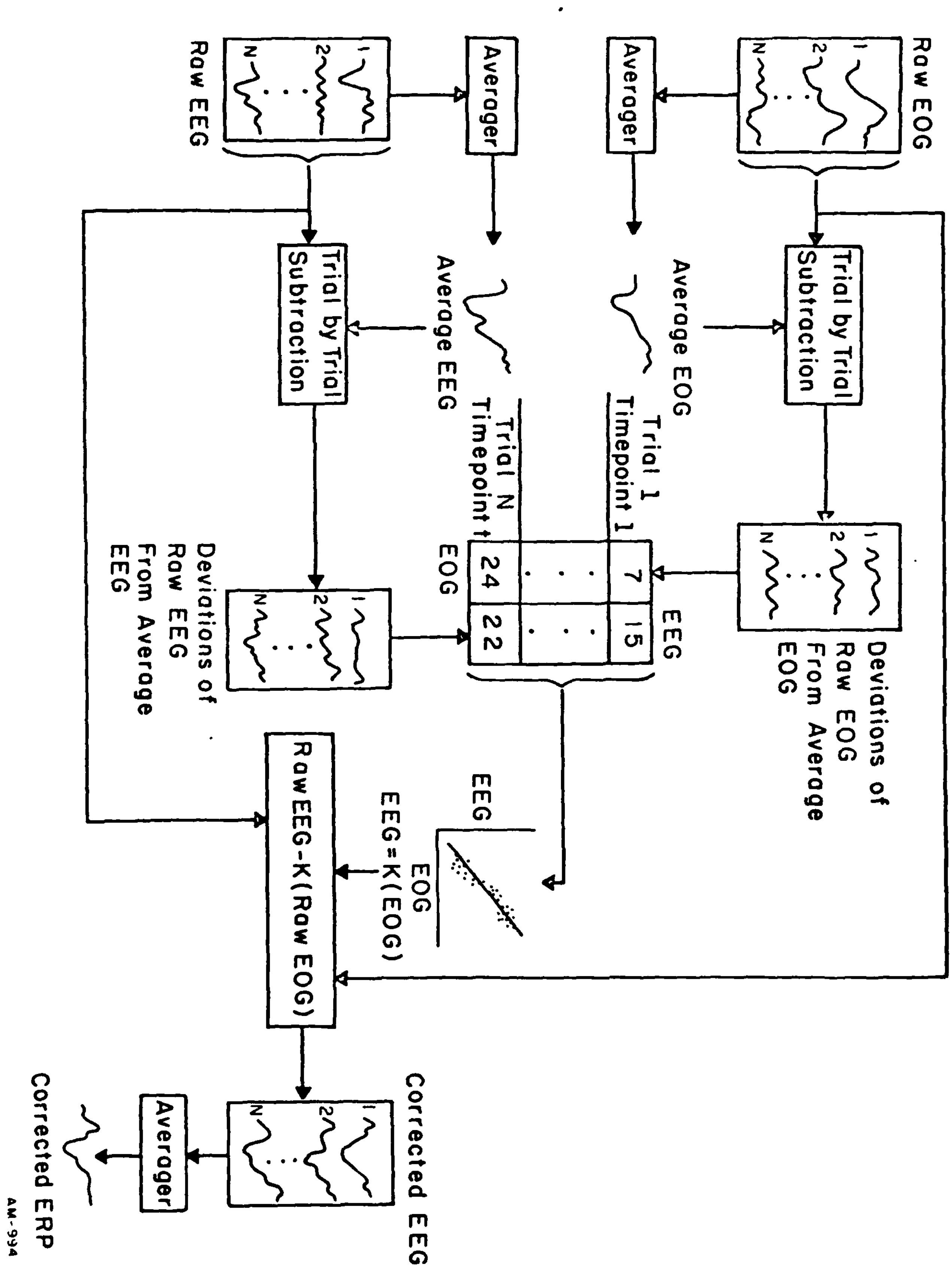


Figure 2



Raw EOG
1
2
N
Averager
Average EOG
Trial by Trial Subtraction
Deviations of Raw EOG From Average EOG
Raw EEG
1
2
N
Averager
Average EEG
Trial by Trial Subtraction
Deviations of Raw EEG From Average EEG
Trial 1 Timepoint 1
Trial N Timepoint t
EOG
EEG
7
15
24
22
EEG
EOG
EEG=K(EOG)
Raw EEG-K(Raw EOG)
Corrected EEG
1
2
N
Averager
Corrected ERP
AM-994

Figure 1

10. Figure Legends

Figure 1. Schematic representation of the Eye Movement Correction Procedure (EMCP) (from Gratton et al, 1983).

Figure 2. Average ERPs sorted by discriminant function classification and type of stimulus. The waveforms represent an average of 16 subjects (reproduced from Squires & Donchin, 1976).

Figure 3. Tree diagram of discriminant scores calculated for ERPs elicited by high and low pitched tones. The discriminant scores are plotted as a function of stimulus sequence (reproduced from Squires et al, 1976).

Figure 4. Plot of four sets of component loadings derived from a principal components analysis of an ERP data set. Each of the component loading vectors is composed of 128 points corresponding to 128 time points (100 Hz digitizing rate) in the waveforms.

Figure 5. Vector Analysis: Geometrical representation of a two-element vector ($\underline{v}$). Values of the corresponding cartesian and polar coordinates are also shown.

Figure 6. Vector Filter: Projection of the observed vector ($\underline{v}$) on the target component vector ($\underline{c}$). Values of the cartesian and polar coordinates for observed, target, and error ($\underline{e}$) vector are also shown.

9. Footnotes

Footnote 1: (Page 94) There is a typographical error in the formula given in Jennings and Wood (1976). The penultimate parenthesis should be deleted.

Footnote 2: (Page 101) The content of Section 5.3 owes much to Cohen and Cohen (1975). As the present chapter is being written, a new edition of their book is in press.

Weerts, T. C., & Lang, P. J. The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). Biological Psychology, 1973, 1, 1-19.

Wiener, N. Extrapolation, interpolation, and smoothing of stationary time series. Cambridge: MIT Press, 1964.

Wilder, J. The law of initial values in neurology and psychiatry: Facts and problems. Journal of Nervous and Mental Disease, 1957, 125, 73-86.

Wilson, K. V. A distribution-free test of analysis of variance hypotheses. Psychological Bulletin, 1956, 53, 96-101.

Wilson, R. S. CARDIVAR: The statistical analysis of heart rate. Psychophysiology, 1974, 11, 76-85.

Winer, B. J. Statistical principles in experimental design, (2nd Ed.) New York: McGraw-Hill, 1971.

Woodward, J. A., & Overall, J. E. Multivariate analysis of variance by multiple regression methods. Psychological Bulletin, 1975, 82, 21-32.

Woody, C. D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical and Biological Engineering, 1967, 5, 539-553.

Woody, C. D., & Nahvi, M. J. Application of optimum linear filter theory to the detection of cortical signals preceding facial movement in the cat. Experimental Brain Research, 1973, 16, 455-465.

Ungar, P., & Basar, E. Comparison of Wiener filtering and selective averaging of evoked potentials. Electroencephalography and Clinical Neurophysiology, 1976, 40, 516-520.

Van Egeren, L. F. Multivariate statistical analysis. Psychophysiology, 1973, 10, 517-532.

Venables, P. H., & Christie, M. J. Electrodermal activity. In, I. Martin & P. H. Venables (Eds.), Techniques in psychophysiology. Chichester, England: Wiley, 1980.

Venables, P. H., & Martin, I. Skin resistance and skin potential. In, P. H. Venables & I. Martin (Eds.), A manual of psychophysiological methods. Amsterdam: North-Holland, 1967a.

Venables, P. H., & Martin, I. (Eds.) A manual of psychophysiological methods. Amsterdam: North-Holland, 1967b.

Walter, D. O. A posteriori "Weiner filtering" of average evoked responses. Electroencephalography and Clinical Neurophysiology, 1968, Supplement 27, 59-70.

Wastell, D. G. Statistical detection of individual evoked responses: an evaluation of Woody's adaptive filter. Electroencephalography and Clinical Neurophysiology, 1977, 42, 835-839.

Wastell, D. G. On the correlated nature of evoked brain activity: Biophysical and statistical considerations. Biological Psychology, 1981, 13, 51-69. (A)

Wastell, D. G. PCA and varimax rotation: Some comments on Roessler and Manzey. Biological Psychology, 1981, 13, 27-29. (B)

Squires, K. C., Wickens, C. D., Squires, N. C., & Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 1976, 193, 1142-1146.

Squires, K. C., Donchin, E., Herning, R. I., & McCarthy, G. On the influence of task relevance and stimulus probability on event-related potential components. Electroencephalography & Clinical Neurophysiology, 1977, 42, 1-14.

Stern, R. M., Ray, W. J., & Davis, C. M. Psychophysiological recording. New York: Oxford University Press, 1980.

Sternbach, R. A. Some relationships among various "dimensions" of autonomic activity. Psychosomatic Medicine, 1960, 22, 430-434.

Stevens, J. P. Comment on Olson: Choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 1979, 86, 355-360.

Stevens, J. P. Power of the multivariate analysis of variance tests. Psychological Bulletin, 1980, 88, 728-737.

Strong, P. Biophysical measurements. Beaverton, Oregon: Tektronix, 1970.

Tatsuoka, M. M. Discriminant analysis: The study of group differences. Champaign, Illinois: Institute for Personality and Ability Testing, 1970.

Tatsuoka, M. M. Multivariate analysis: Techniques for educational and psychological research. New York: John Wiley & Sons, Inc., 1971.

Thurstone, L. L. The vectors of mind. Chicago: The University of Chicago Press, 1935.

Roessler, F., & Manzey, D. Principal components and varimax-rotated components in event-related potential research: Some remarks on their interpretation. Biological Psychology, 1981, 13, 3-26.

Ruchkin, D. S., & Glaser, E. M. Simple digital filters for examining CNV and P300 on a single-trial basis. In D. Otto (Ed.), Multidisciplinary perspectives in event-related potential research. Washington, D.C.: U.S. Government Printing Office, 1978.

Ruchkin, D. S., Sutton, S., & Stega, M. Emitted P300 and slow wave event-related potentials in guessing and detection tasks. Electroencephalography & Clinical Neurophysiology, 1980, 49, 1-14.

Schnore, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. Journal of Experimental Psychology, 1959, 58, 117-128.

Sencaj, R. W., Aunon, J. I., & McGillem, C. D. Discrimination among visual stimuli by classifications of their single evoked potentials. Medical and Biological Engineering and Computing, 1979, 17, 391-396.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Simons, R. F., Ohman, A., & Lang, P. J. Anticipation and response set: Cortical, cardiac, and electrodermal correlates. Psychophysiology, 1979, 16, 222-233.

Squires, K. C., & Donchin, E. Beyond averaging: The use of discriminant functions to recognize event-related potentials elicited by single auditory stimuli. Electroencephalography & Clinical Neurophysiology, 1976, 41, 449-459.

Overton, D. A., & Shagass, C. Distribution of eye movement and eye blink potentials over the scalp. Electroencephalography and Clinical Neurophysiology, 1969, 27 , 544-549.

Picton, T. W., & Stuss, D. T. The component structure of human event-related potentials. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, motor and sensory processes of the brain: Electrical potentials, behavioral and clinical use. Netherlands: North Holland Biomedical Press, 1980.

Porges, S. W. Developmental designs for infancy research. In J. D. Osofsky (Ed.), Handbook of infant development. New York: Wiley, 1979.

Porges, S. W., Bohrer, R. E., Cheung, M. N., Drasgow, F., McCabe, P. M., & Keren, G. New time-series statistic for detecting rhythmic co-occurrence in the frequency domain: The weighted coherence and its application to psychophysiological research. Psychological Bulletin, 1980, 88, 580-587.

Press, S. J. Applied multivariate analysis. New York: Holt, Rinehart & Winston, 1972.

Ragot, R. A., & Remond, A., EEG field mapping. Electroencephalography and Clinical Neurophysiology, 1978, 45 , 417-421.

Remond, A. Construction et ajustment des enregistrement cartographiques et spatio-temporals des EEG. Revue Neurologique, 1962, 106, 135-136.

Richards, J. E. Multivariate analysis of variance of repeated physiological measures. Paper presented at the Annual Convention of the Society for Psychophysiological Research, Vancouver, B. C., October, 1980. (Abstract: Analyzing repeated physiological measures with multivariate ANOVA.) Psychophysiology, 1981, 18, 148.

Mulaik, S. A. The Foundations of factor analysis. New York: McGraw Hill, 1972.

Myers, J. L. Fundamentals of experimental design, (3rd. Ed.). Boston: Allyn and Bacon, 1979.

Myers, J. L., DiCecco, J. V., White, J. B., & Borden, V. M. Repeated measures on dichotomous variables: Q and F tests. Psychological Bulletin, 1982, 92, 517-525.

Nahvi, M. J., Woody, C. D., Ungar, R., & Sharafat, A. R. Detection of neuroelectric signals from multiple data channels by optimum linear filter method. Electroencephalography & Clinical Neurophysiology, 1975, 38, 191-198.

Naitoh, P., & Sunderman, S. Before averaging: Preprocessing slow potential data with a Wiener filter. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research, EPA-600/9-77-043. Washington, D.C.: U.S. Government Printing Office, 1978, pp. 573-578.

Nicewander, W. A., & Price, J. M. Dependent variable reliability and the power of significance tests. Psychological Bulletin, 1978, 85, 405-409.

Olson, C. L. On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 1976, 83, 579-586.

Olson, C. L. Practical considerations in choosing a MANOVA test statistic: A rejoinder to Stevens. Psychological Bulletin, 1979, 86, 1350-1352.

Overall, J. E., & Woodward, J. A. Nonrandom assignment and the analysis of covariance. Psychological Bulletin, 1977, 84, 586-594.

WORKLOAD--AN EXAMINATION OF THE CONCEPT

By

Daniel Gopher
Faculty of Industrial Engineering
Israel Institute of Technology
The Technion
Haifa, Israel
011 972 4 292 130

and

Emanuel Donchin
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, Illinois 61820
(217) 333-0632

Preparation of this paper, during the long period of its gestation, has been aided by several organizations and many colleagues. Lloyd Kaufman, Ken Boff, Jack Beatty, Chris Wickens and Sandra Hart have commented on different versions of the paper. Maya Weil and Nira Arazi at the Technion, and Sharon Cummings at the University of Illinois helped with the preparation of the Bibliography. Barbara Mullins prepared the manuscript, and Barbara Hartman and Alfreda Mitchell aided in a variety of ways. The authors' research programs have been aided by AFOSR, ONR, and DARPA. Additional support was provided by NASA Ames Research Center, The Man-Vehicle Research Division, under Grant number NAGW-494 to Daniel Gopher. We appreciate Lloyd Kaufman's patience as this paper was being written transatlantically. Parts of Section 3.9 are revisions of material presented by Donchin, Kramer, and Wickens (1982).

E. Donchin is particularly appreciative of the office provided at NYU which allowed him to escape the distractions of Champaign-Urbana and concentrate on writing in the serene atmosphere of Washington Square. The authors have shared the effort and the responsibility in the preparation of this chapter. The order of authorship was determined by the toss of a coin.

To appear in K. Boff & L. Kaufman (Eds.), Handbook of Perception and Human Performance. New York: Wiley & Sons, in press.

TABLE OF CONTENTS

		Page
1.	INTRODUCTION	1
1.1	Task Difficulty and Workload	1
1.2	Workload and the Limitations on Performance	5
1.3	The Logical Status of Workload	9
1.3.1	A Definition of Workload	9
1.3.2	Workload as an Intervening Variable	11
1.3.3	Workload as a Hypothetical Construct	14
1.4	Workload and Attention	15
2.	THE LIMITATIONS OF THE CENTRAL PROCESSOR: HISTORICAL BACKGROUND	17
2.1	Origins of the Notion of a Central Limited Processor	17
2.1.1	The Human as an Information Channel	21
2.1.2	The Putative Advantages of Information Theory	24
2.2	Channel Capacity as a Measure of Limits on the Central Processor	25
2.2.1	Absolute Judgment	27
2.2.2	The Span of Short-Term Memory	29
2.3	Summary	34
2.4	Bottleneck Models of Human Limitations	35
2.4.1	Single Bottleneck Models	37
2.4.2	Welford's Single-Channel Model	41
2.4.3	The Breakdown of the Single Channel	48

2.5	The Architecture of the Central Processor...............	56
2.6	Energy Constraints on Processing Capabilities...........	62
2.6.1	The Single Resource Model............................	62
2.6.2	Multiple Resource Models.............................	70
2.7	Controlled and Automatic Processes.....................	77
2.8	The Nature of Capacity Limitations.....................	80
2.8.1	Dimensions of a Load Profile.........................	81
2.8.2	Conscious Control and Allocation Policy..............	83
3.	TECHNIQUES FOR THE MEASUREMENT OF WORKLOAD..............	88
3.1	Criteria for the Evaluation of Workload Measures........	88
3.2	Workload as a Property of the Operator/Task Loop........	89
3.2.1	Normative and Descriptive Approaches.................	90
3.2.2	Overview of Section..................................	93
3.3	Subjective Measures....................................	95
3.3.1	Consistency of Subjective Estimates..................	99
3.3.2	Theoretical Considerations in the Interpretation and Use of Subjective Measures of Workload.............	103
3.3.3	Methodological Considerations........................	106
3.4	Performance Measures, Primary Task.....................	107
3.5	Arousal Measures.......................................	111
3.6	Specific Measures......................................	113
3.6.1	Selection of a Secondary Task........................	115
3.6.2	Types of Interference and Lack of Interference.......	118
3.6.3	The Problems of Concurrency..........................	123

3.6.4	Allocation Policy	127
3.7	Performance Operating Characteristics	130
3.8	Basic Assumptions of the POC Methodology	140
3.9	Psychophysiological Measures of Workload	142
3.9.1	Introductory Comments on the P300 Component	143
3.9.2	The Latency of the P300 Component	147
3.9.3	P300 and Perceptual/Central Processing Resources	153
3.9.4	P300 and Resource Reciprocity	163
3.9.5	Summary and Conclusions	167
4.	EPILOGUE	167
5.	REFERENCES	174

1. INTRODUCTION

1.1 Task Difficulty and Workload

The study of measurement of mental workload absorbs substantial energy, resources, and effort. Major conferences devoted to the analysis and measurement of workload (Moray, 1979; Frazier & Crombie, 1982) follow a fairly uniform course. The concept of workload is examined, attempts at a definition are made, and the usual conclusion is that workload is a multidimensional, multifaceted concept that is difficult to define. It is generally agreed that attempts to measure workload relying on a single representative measure are unlikely to be of use.

The conferees generally proceed to examine many different procedures to measure and analyze workload. Practitioners tend to imply that the measure, or class of measures, they advocate is uniquely preferable to many of the other proposed measures. Persuasive arguments are presented for the unique suitability of (a) subjective measures of workload (Sheridan, 1980), (b) secondary task procedures (Jex, 1976; Pew, 1979) and (c) physiological measures (Donchin, 1979; Mulder, 1979). These writers are actually quite correct in holding to their particular views, since the measures described often yield an interesting set of relations and in some circumstances seem to meet the practical needs of design engineers.

O'Donnell and Eggemeier (Chapter 42) provide a survey, written from the perspective of the design engineer, organizing the many techniques proposed for measuring workload. As they indicate in the opening paragraphs of their chapter, the interest in workload arises in evaluating "task difficulty." The assessment of workload is a direct assessment of a class of difficulties which operators confront when performing an assigned task. The need for this concept, and for the vast literature devoted to its analysis and measurement, is generated by the complexity of the deceptively simple concept "difficulty." That tasks vary in their difficulty is, of course, obvious. It is easier to read Agatha Christie than James Joyce; it is more difficult to fly an airplane than to operate a washing machine. It is also clear that the same task will prove more difficult to some individuals than to others. Furthermore, the ease with which the same individual performs a given task on different occasions may vary. Yet, despite the apparent simplicity of the concept and the fairly easy judgments observers can make regarding the difficulty of tasks the measurement of task difficulty is in itself a rather difficult task.

The problem exists because the difficulty of any task cannot be inferred directly from its physical (or "structural") description, but rather from the interaction

between task and operator. Hence, system designers need to be able to discriminate among alternate design options in favor of those which will ease the operator's task. On occasion, system managers must be able to monitor variations in the difficulty a task presents to an operator actively using a system. It is on such occasions that the need to measure "task difficulty" arises. And it is on such occasions that the apparent simplicity of the measurement task is revealed as horrendously complex.

Superficially the measurement task appears quite straightforward. One simply asks an operator if the task is difficult and uses subjective response as an index of difficulty. As discussed in Section 3.3 (and as O'Donnell and Eggemeier note in Chapter 00) such subjective reports have serious limitations. An operator is often an unreliable and invalid measuring instrument. Neither can it be assumed that the quality of performance is a good measure of the difficulty of a task. People often cope with increases in task difficulty by increasing mental and physical effort devoted to the task, so that performance may remain stable despite a great increase in difficulty. The measurement of difficulty must capture the interaction between these relevant variables. "workload" is the label assigned to this interactive feature.

An analogy might clarify the matter. Consider a simple resistive circuit in which a voltage source, say a battery, is imposing a voltage across some resistor. Both the voltage supplied by the battery and the resistance that characterizes the resistor depend on the properties of these devices regardless, to a first approximation, of the circuit in which they are embedded (though both voltage and resistance depend on such circumstances as the ambient temperature). However, specifying the voltage or the resistance by themselves tell us very little about the circuit. There is, however, another variable that actually captures the properties of the circuit, reflecting the <u>interaction</u> between the voltage and the resistance. The current flowing in the circuit is determined jointly by the properties of the battery and the resistor and is therefore a parameter of the circuit, rather than of its individual elements. Variations in the current, given fixed voltage and resistance, can also be used as a means for assessing the degree to which ambient conditions affect the properties of circuit elements.

We suggest that Workload is analogous to Current. Of course, unlike Current we cannot derive the Workload associated with the interaction between a specific operator and a specific task from an equation that combines measureable attributes of the operator with measureable attributes of the

task to yield workload. The intent of this chapter is to review the search for relevant task and subject attributes as well as the combination rules that may be applied to these measurable variables so as to provide a measure of Workload. However, it is important to emphasize at the outset that we view Workload as an attribute of the interaction between a person and a task. The concept is introduced, we shall argue, as a hypothetical construct to summarize the difficulty that a task presents to an operator. This review will therefore be concerned with two main topics: (a) the manner in which the task-operator loop can be modeled to derive useful explications of the concept of workload, and (b) the theoretical status of the different modes of measurement employed in the assessment of workload.

1.2 Workload and the Limitations on Performance

The term "workload" is used to describe aspects of the interaction between an operator and an assigned task. "Tasks" are specified in terms of their structural properties; a set of stimuli and responses are specified with a set of rules that map responses to stimuli. There are, in addition, "expectations" regarding the quality of the performance, which derive from knowledge of the relation between the structure of the task and the nature of human capacities and

skills. Expectations may also be based on the individual operator's past performance or on knowledge of the way others perform similar tasks. These expectations are frequently not met even though the individual is motivated to accept the assignment and intends to perform according to expectations. Often we ascribe such failures in performance to increased difficulty of the task. In the attempt to explain and cope with these interactions the concept of workload finds its primary use.

Often a failure to perform is easy to explain as the task was evidently beyond the capacities of the organism. If one required a human to jump from one rim of the Grand Canyon to the other it is very unlikely that the concept of Workload would be invoked to explain an inability to perform the task. One would likely say that this task is beyond human ability. In the terminology of the previous section, we do not attribute the failure to an interaction between the operator and the task. Similarly, it is not particularly useful to ascribe to excessive workload the inability of a monolingual English speaking student to read yesterday's conversation exercise in Latin. The teacher will, no doubt, assume that the failure in performance indicates that the student has never learned, and therefore lacks the capacity, to perform the task. In other words, there are many occasions in which a

failure to perform according to expectations is attributed to very obvious limitations on the capacities, either innate or learned, of the organism rather than to the interaction between the organism and the task. In such cases there is no need to invoke workload as an explanatory concept.

However, often the limitations are not obvious. If the student has read the exercise many times in the past, and is generally known to be attentive in class and eager to please, a sudden failure to perform will force the teacher to assume that "other factors" have intervened to limit performance. Let us assume, for simplicity, that motivation is never a problem in the context of our discussion. The intention to perform is genuine, as is the deep desire to perform well. If this is clear, and if performance shows a decrement despite a demonstrated past ability to perform adequately, then we are inclined to interpret a failure to perform according to expectations as evidence for a deterioration in the performer's ability to execute the task that may be due to an increase in workload.

The deterioration may be due to pathology. A stroke may have incapacitated the subject. In such an event we shall again ascribe the difficulty to the pathology without feeling bound to invoke the concept of workload. On the other hand, if we know that there has been no overt pathology, then we may

indeed suggest that, for reasons still to be determined, the demands imposed on the operator by the task have become larger than they were when performance was at its best. It is at this theoretical juncture that the concept of workload is normally introduced. That is, workload is invoked to account for those aspects of the interaction between a person and a task that cause task demands to exceed the person's capacity to deliver. Note that the concept is needed only for those cases in which the required performance of a specified quality is clearly within the performer's current repertoire. In this discussion we are ignoring, in keeping with the tradition in discussions of workload, the effects of learning on an individual's repertoire. We are concerned with the deployment of the available responses at the time a task was assigned without considering how the repertoire has been created. Subsequent traini 'g may well change the repertoire. Indeed, in Section 2.7 we do discuss the effects of practice on workload, but restrict our interest to improvement in performance of a previously acquired skill rather than on the acquisition of new skills.

Workload seems to refer to a cost the operator incurs as tasks are performed. In principle, if a capacity to perform is available, a failure to perform must imply limits to its use. In the same way that physical workload was used by

Ergonomists to specify limits on the muscles' ability to deliver, so is mental workload invoked to explain the mind's inability to deliver. Because muscular work is directly observable, its physiology and mechanisms are relatively well understood. But mental workload is clearly an attribute of the information processing and control systems that mediate between stimuli, rules, and responses. Mental workload is an attribute of the person-task loop, and the effects of workload on human performance can therefore be examined only in relation to a model of human information processing. The logical status of the concept depends on one's stance regarding cognition. The following section examines mental workload as a concept within this framework.

1.3 The Logical Status of Workload

1.3.1 A Definition of Workload

The current concept of workload implies that limitations exist in the information processing structures, making it difficult for a person to fully use the information processing apparatus in the service of the target task. The key assumption is that to perform a task the organism uses effectors through which the overt responses are made as well as sensors through which information is gathered. The path

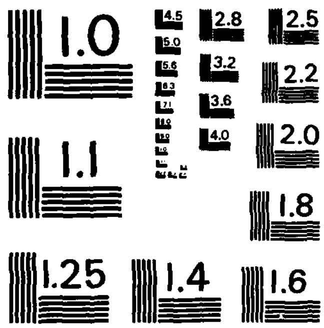
between sensors and effectors is an elaborate information-processing apparatus, with structural properties and a limited capacity. These limitations on the capacity of the information-processing system must be measured and modeled if we are to account for performance failures attributed to mental workload.

In other words, mental workload may be viewed as the difference between capacities of the information-processing system that are required for task performance to satisfy performance expectations and the capacity that is available at any given time. Task difficulty is thus manifested by a difference between the expected and the actual performance. It is necessary to specify who is doing the expecting if the term "Expected Performance" is to have precise meaning. The level of "expected" task performance in any constellation is established by the level of performance of the same task under the least demanding circumstances. For example, if a person can listen to two conversations when they are presented singly, a failure to monitor both concurrently requires the invocation of workload.

This definition implies that workload is an intervening variable rather than a hypothetical construct. MacCorquadale and Meehl (1948) discussed the distinction between these two

AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 3/9

UNCLASSIFIED 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

meta-theoretical terms. An intervening variable is a theoretical concept, such as electrical resistance, that is,

> . . . simply a quantity obtained by a specified manipulation of the values of empirical variables; it will involve no hypothesis as to the existence of unobserved entities or the occurrence of unobserved processes; it will contain, in its complete statement for all purposes of theory and prediction no words which are not defined either explicitly or by reduction sentences in terms of the empirical variables (MacCorquadale & Meehl, 1948, pgs. 95-107).

1.3.2 Workload as an Intervening Variable

If we could elucidate functions that relate observed decrements in performance to the specified criterial performance, these functions would define workload in the same way that resistance is defined by Ohm's Law. The system's inability to perform as desired would be attributed to some internal property labeled Workload that resists the execution of the task as specified. Of course, in the same way that current could be further analyzed and explained in terms of atomic theory, so we would ultimately expect to account for workload in cognitive and neurophysiological terms. But, even

without this clarification the concept can be employed usefully in theories of human performance and in applications of these theories in Engineering Psychology.

However, the situation is a bit more complex. The fact is that on too many occasions the observations that compel us to invoke the concept of workload are inconsistent with its perception as an intervening variable. Quite clearly it is often virtually impossible to infer the degree to which a given task is beyond the capacity of a given individual on any occasion. Operators are seen to cope with extraordinary demands so that their observed performance displays no obvious variance. To define workload by direct observations on performance, we must conclude that there are no changes in workload even though our common sense and subsequent analysis suggest otherwise.

Consider, for example, a human presented with a plank which is 1 foot wide and 20 feet long. It is solid and sturdy and can, beyond doubt, carry the person's weight without breaking. Assign the person the task of getting on the plank at its proximal end and walking to its distal end. Measuring task performance in terms of the person's ability to walk the plank, score "1" for a successful traverse and "0" for a failure. Be even more careful and measure the time interval between the instant the plank is approached to the instant the

task is completed. It is conceivable that for an experienced person, walking the plank will be achieved with equal facility when the plank is laid on the floor, when it is suspended across two chairs so that it is 3 feet from the ground, and when it is suspended above the middle ring of a circus at 50 feet. Performance, as we defined it, may show absolutely no variance. And yet, we all know intuitively that the workload associated with each of the three tasks is quite different.

We bother to note that workloads are different because we would like to know how much additional work we can impose on the plank-walker. We are persuaded that the person walking the plank as it is laid on the floor can undertake additional tasks, while the one hovering over the rink will resist undertaking more. But what is it about walking a plank when it is very high that is different from walking it when it is lying on the floor? Of course, in one case the consequences of failure are more disastrous than in the other. The increased danger will "concentrate the mind" as Samuel Johnson said of the condemned. We would not be surprised that when balancing above a precipice the walker is less likely to conduct a conversation, and seems oblivious of all surrounding activity. We may also note that task performance is far more robust to changing conditions when the plank is at zero height than when it is up high. As we continue this description we

invoke observables that reflect differences in task performance. However, these observables were not included in our initial definition of the task. Neither are we as sure when we invoke them that they will indeed catch the flavor of the difference we are sensing. However, we allude precisely to the effect that raising the plank has on task performance when we refer to mental workload.

1.3.3 Workload as a Hypothetical Construct

Workload logically carries the excess meaning that, according to MacCorquadale and Meehl (1948), characterizes hypothetical constructs. This label is applied to concepts that . . . "involve terms which are not wholly reducible to empirical terms; they refer to processes or entities that are not directly observable (although they need not be in principle unobservable)".

In this framework, Workload is a concept like Electron, rather than like Current. We imply that we are describing some entity or some property of entities that is not given entirely by the relationship between our empirical observations. At the same time we assume that this excess meaning can be captured, studied, and measured in ways that would advance our understanding of the system and make it possible to use the concept for practical activities.

1.4 Workload and Attention

By defining workload in terms of the limitations on the capacity of an information processing system we are underlining the close affinity between the literature concerned with workload and the literature that focuses on attention. In both bodies of literature the prime concern of investigators is to assess, and possibly explain, performance limitations that are manifested despite an apparent ability of the individual to perform a task. Thus, the psychologist attending the proverbial cocktail party (Cherry, 1957), and listening to at least two concurrent conversations is clearly capable of following either. Yet, one conversation may come to predominate at the expense of the other. In fact, the content of one of the conversations is often ignored. This failure to converse is remarkable, and is related to our discussion of workload, because the person is clearly capable of switching "attention" from one conversation to another depending on the level of interest the ignored conversation promises.

This ability to switch among tasks implies that the contending tasks are all within the person's repertoire when performed singly. The limitation, as in the case of workload, appears to be in the system's capacity to deal with multiple demands. As with workload, investigators have suggested that

the limit on attention reflects constraints inherent in the structure and organization of a central processor. Our discussion of workload begins by examining various attempts to use the concept of a central processor and its limitations on the capacity of such a processor in modeling human performance in a number of domains.

It is useful to explicate the concept of a limited capacity processor at the beginning of this discussion since most of the methods proposed, and employed, in the study of workload can be viewed as attempts to identify and quantify these limitations. Following the discussion of the limited capacity processor we shall review the same classes of measures of workload discussed in detail by O'Donnell and Eggemeier (Chapter 42). We will first consider the introspective, or "subjective" measures (Moray, 1982), which assume that operators can perceive their own limitations and make truthful, or at least usable, reports. The second class of procedures assumes that the operators' introspections need to be augmented by additional objective observations. Secondary task techniques (Ogden, Levine, & Eisner, 1979) assume that it is possible to assess the limitations on capacity by imposing yet another task on the subject. Failures to perform this secondary task are taken as evidence that the "primary" task is exceeding the limits on the

system's capacity. A final technique, the psychophysiological, records the activity of certain bodily systems and assumes, as do the subjective measures, that an individual's body can reveal the load on the system. Yet, it replaces introspection by a reliance on the wisdom of the body (Cannon, 1932).

Each of these approaches is illustrated and the advantages and drawbacks within the framework of our concept of Workload are considered. The final section proposes an approach to the analysis and measurement of workload that integrates these considerations.

2. THE LIMITATIONS OF THE CENTRAL PROCESSOR: HISTORICAL BACKGROUND

2.1 Origins of the Notion of a Central Limited Processor

The notion of a central limited processor can be traced to the concept of attention discussed by the pioneers of scientific psychology, such as Williams James (1890) and Titchner (1908). Another important influence has come from communication engineering in the years following World War II (Broadbent, 1958; Miller, 1956). William James and his contemporaries equated selective processing and attention with the mechanism that regulates consciousness. James's discussion of the limits of consciousness implied, in modern

terms, that the mechanisms operate as if they were a central limited processor. It is so structured that it can attend to one single event at any one time. In James's words:

> . . . Every one knows what attention is. It is the taking possession by the mind, in a clear and vivid form of one out of what seem several simultaneously possible objects or trains of thought. Focalization concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatter-brained state which in French is called DISTRACTION, and ZERSTREUTHEIT in German (pp. 403-404).

When making these claims, James was careful to specify his notion of the possible nature of such "single" events that can capture the totality of consciousness at a given moment:

> . . . The number of things we may attend to is altogether indefinite, depending on the power of the individual intellect, on the form of the apprehension, and on what the things are. When apprehended conceptually as a connected system, their number may be large. But

> however numerous the things, they can only be known in a single pulse of consciousness for which they form one complex 'object', so that properly speaking there is before the mind at no time a plurality of IDEAS, properly so called. (p. 405)

This definition of a processing event anticipates by many years the contemporary distinction between isolated items and chunks of information in working memory (Miller, 1956), Gestalt and grouping principles in perception, integrality of dimensions (Garner, 1974), and effect of reorganization and training (Neisser, 1976; Schneider & Shiffrin, 1977). Note that James and his contemporaries tended to identify attention with the contents of consciousness. The limitations on the central processor are, in this framework, identical to the limitations on consciousness. However, an exclusive identification of attention and of information processing with the operation of consciousness is not tenable. It may do when one attempts, as James did, to develop a comprehensive catalog of mental activities. However, when the goal is a system to predict behavior it can not be ignored that the scope of information processing far exceeds the scope of consciousness. Indeed, it is possible to assert that many of the processes of key interest to the theorist of Human Information Processing

are not, and cannot be, accessible to consciousness (Broadbent, 1982; Kaufman, 1979; Kahneman & Treisman, 1983; Posner & McCleod, 1983).

Therefore, the phenomena of attention and workload encompass a broad spectrum of selection, transformation, and processing activities only part of which is accessible to consciousness. Thus, with mounting evidence on the structure of the information processing system, and with increasingly detailed functional descriptions of the systems, it becomes evident that theories of attention and workload account for a vast array of processes about which no subjective information is available. The limitations imposed on performance may derive from mechanisms that are opaque to subjective assessment, such as memory search, the construction of grammatical sentences, and allocation of attention to one of several channels. It is unfortunate and counterintuitive that consciousness reveals only a smattering of the workings of the system of interest. Yet, if we are to model and measure limits on the information processing mechanisms we must go beyond William James and quantify limitations outside of consciousness. It will be easier to review the nature of such limitations after we introduce the concept of "information channel" that has played an important role in the analysis of attention and workload.

2.1.1 The Human as an Information Channel

The effects of communication engineering on psychological theory can be seen by examining the effect of information theory on experimental psychology during the 1950's and 1960's. The analysis is of particular interest here because the very notion of a "limited capacity channel" that is central to current discussions of workload derives from communication engineering and in particular from the Theory of Information developed by Shannon and Weaver (1949). The major contribution of Information Theory to psychology has been the acceptance of "information" as a commodity that can be manipulated, transmitted, and transformed. Moreover, discussions of information could proceed without any reference to the physical implementations of the information processing device. One can describe the properties of information flow, or of the transformation of information, by the same formal system whether the system is implemented by vacuum tubes, transistors or, it is hoped, neurons. Thus psychologists attempted to explain human information processing in terms of the flow of information within the organism. The human processing system was likened to a communication channel that processes messages and transmits information from a stimulus set to a response set (Attneave, 1959).

In ordinary usage, the term "information" refers to the properties of events, stimuli, or messages that reduce one's uncertainty about the true state of affairs. In formal information theory, information is "transmitted" to the extent that appearance of a message reduces the prior probability of a response in the response set. The greater the change between the prior and posterior probability of the response, as a result of the presentation of a message, the greater the amount of information transmitted by the message.

A key concept in Information Theory is that of the Communication Channel, which exists between any two communicating points. It is defined by its capacity to transmit information between sender and receiver and it is characterized by a number of quantifiable parameters, the most crucial of which, for our discussion, is that of channel "capacity." One of Shannon and Weaver's (1949) principal insights was that channels can vary in their capacities and that these differences can be quantified within the framework of Information Theory. A channel displays its full capacity if it imposes no reduction in the transmission of information between sender and receiver. Degradations of channel capacity are measured as decrements in information transmission. If the uncertainty of a receiver is reduced at a lower rate than would result from information available from the sender,

Attempts to develop models of the limited processor in terms of strict Information Theory were abandoned by the late 50's in favor of models postulating internal mechanisms that operate on information and determine channel capacity. While it took as its starting point a metric of capacity, the new models were based on one or another mechanism that was responsible for the observed limitations in performance. Investigators in this phase were not concerned with precise metrics for expressing the limits, but rather with demonstrating empirically that their structural model of the mind was valid. The emphasis was on postulating and demonstrating bottlenecks in information flow. The workload construct played a minimal role in these efforts as they were primarily viewed by the practitioners as studies of "attention." Yet, because this work provided a framework within which the workload construct developed, we will review briefly a couple of prominent "bottleneck" models and trace their influence on the development of workload as a hypothetical construct. In reviewing these models we will distinguish between single and multiple bottleneck models. The former assume that the limitation on performance can be localized in one universal mechanism; the latter take a pluralistic view of human limitations, admitting that we can be imperfect in many different ways.

2.4 Bottleneck Models of Human Limitations

An important consequence of the view of the human as a processor with an upper limit on information transfer was the developing recognition that different tasks impose different demands on this processor and therefore "load" it to different extents. Thus, the construct workload developed naturally within the framework of the Information Theory <u>zeitgeist</u> of the 50's. Of even greater consequence was the concept of a "left-over" or "spare" capacity. If capacity is measurable, and if different tasks consume different amounts of capacity, then some tasks consume less than full capacity and must leave a residual of spare capacity, a hypothetical quantity that might be measurable. The reader may recognize in these questions issues that accompany any discussion of mental workload. We will return later to a detailed discussion of this complex topic. Note however that within the framework of Information Theory-based analysis, these issues played a central role and were associated with a formal technical meaning. If capacity is limited to about 2.5 to 3 bits of information per sec, and the demands imposed by a certain task are 2 bits/sec, the processor can be said to have 1 bit of spare capacity. Although the calculus was too simple as a model of reality the underlying concept is remarkably viable.

necessary nor sufficient to explain the experimental data. Contemporary models of motor control are more often based on models of the underlying control processes or on the structure of the information processing system underlying movement than they are on Information Theory specification of the movement control channel (Jagacinski, 1980; Keele, 1981).

2.3 Summary

Attempts to use Information Theory to model the limitations on human performance have profoundly influenced models of mental workload to be discussed in the remainder of this chapter. In fact, the construct of limited processing capacity can be traced to Information Theory. The research programs reviewed in the previous section served to demonstrate the applicability, however limited, of the communication channel metaphor and of the tools provided by information theory, to problems in the domain of human performance. Even though it is necessary to expand the concept and to incorporate models of information processing that can not be modeled by the Shannon-Weaver calculus, the effects of Information Theory on the cognitive sciences have been salutary.

appeared to have considerable generality. For example, Hick (1952) has shown that Information Theory can be used in the analysis of speed-accuracy tradeoffs. At the same time, however it became evident that despite its promise, this framework could not deal with such critical determinants of the speed with which choices can be made as the stimulus-response compatibility or the level of practice. Again, the information theory approach encountered difficulties that forced investigators to go, in Bruner's (19??) excellent phrase, "beyond the information given."

Similarly, Fitts's Law provides a good summary of much empirical data in Information Theory terms. Yet, it must be augmented as a description of the control of human movement. Fitts (1954) reasoned that the execution and supervision of movements should obey the same rules of other information processing tasks. He argued that the amount of information contained in the conduct of a specific movement should be a function of the distance (amplitude) and the accuracy of the movement. His "law" in its latest form states that $MT=\log A/(B/2)$, where MT is movement time, A is the amplitude or distance of movement, and B is the width of the target. Fitts's Law has been validated extensively (see Keele, 1981, & Chapter 30, for a review). However, it was shown repeatedly that the rationale based on Information Theory is neither

with the introduction of chunking, the discussion of the limits on human performance abandons the simple elegance of Information Theory and faces the realities of the human information processing system.

The concept of chunking preserved Information Theory as an analytical framework but lost its promise as a tool for uniformly analyzing the processing demands of all tasks. It is impossible to base an analysis of tasks on an examination of the formal properties of the external structure of tasks. Although weakened, the metaphor of the communication channel remains viable, but the attention of researchers has turned to the structure of the human information processing system and the processes that govern its operations. Illustrations of the promise, and the ensuing complications, of the Information Theory model abound. The work on Choice Reaction Time, most notably that of Hick (1952) and Hyman (1953) and Fitts's (1954) studies of motor control, deserve mention.

Hick and Hyman adopted the channel capacity concept to account for the finding that response time increases linearly as a function of the information value of the stimulus expressed in bits per sec. The Hick-Hyman law asserts that $RT= a+bI(x)$; where RT is the reaction time, a and b are parameters and $I(x)$ is the information value of in bits (see Chapter 39, by Wickens, this volume). The Hick-Hyman law

Miller's "magical number 7 plus or minus 2." Thus, the issue seems to be the base on which information measures are calculated. The same physical list may contain items carrying different amounts of information. This has been the basis for Miller's (1956) introduction of the notion of a "chunk." Chunks were described as composite units that result from grouping, organizing, or recoding of a group of otherwise isolated elements (e.g., grouping letters into words, or using mnemonics to memorize a string of digits). The chunk, rather than the list element, becomes the basis for computations of channel capacity. The contradiction between the span of immediate memory and the scope of absolute judgment may be due to a dependence of the span of memory on the number of chunks saved and recalled rather than on the number of individual items presented.

Of course, the introduction of chunking reduces the adequacy of Information Theory as a tool for modeling the limitations on human performance. It is no longer sufficient to specify information in bits per second on the basis of physical properties of the stimuli or the structural properties of the task. If chunking is critical, we must know how chunking operates in each context. What rules govern the process of the reorganization of items, and what are the costs? Indeed, what are the limitations on chunking? Thus,

recall as many items as they can from this list. For example, Hayes (discussed in Miller, 1956) read aloud to subjects five lists, at a rate of one item per sec. Varied in complexity, each list contained either binary digits, decimal digits, or letters of the alphabet. One additional list contained both letters and decimal digits. Finally, one list contained monosyllabic words taken from a vocabulary of 1000 words. Within the framework of Information Theory the amount of information per item in each of the five lists is considerably different. For example, decimal digits carry about 3.3 bits each while isolated English words carry about 10 bits of information per word. Hence, if memory tasks are executed by the same communication channel postulated to underlie the performance of subjects in Pollack's absolute judgment tasks, recall on the different lists should vary accordingly. Yet, subjects were able to recall about 10 items regardless of the list used. Similar results were obtained in other studies (e.g., Pollack, 1953).

Even though these results could not be accommodated by the model used to account for the absolute judgment tasks, the mechanism of storage and retrieval encountered limitation at a remarkably similar point. If one considers strictly the number of items in the list, regardless of their identity, recall converges again to 2.5 bits per list, in accord with

The evident success of this enterprise was deceptive. Information Theory indeed served well to model the fairly simple situations to which it was applied in the early 50's. However, in a pattern that will repeat throughout this section, it has become clear that many aspects of the way subjects actually handle absolute judgment tasks display features that can not be described adequately in Information Theory terms. Thus, for example, a readily observed "anchor effect" provides treacherous grounds for Information Theory because it establishes that the judgments subjects make depend on expectations of the context in which stimuli are presented. The probability of events, defined a priori, can not serve as a metric with which to predict performance, or workload, associated with different stimulus sets. It is possible to expand the model to include such considerations, but the elegant simplicity of the basic model is lost.

2.2.2 The Span of Short-Term Memory

Another difficulty with viewing the human information processing system as a classic communication channel arises in the attempt to generalize the conclusions of one application to another. Consider, for example, studies of the span of short-term memory. In these tasks, subjects are presented with a list of items in rapid succession and are asked to

Subjects could identify tones, but not if the number was too large. Thus, the interaction between the subject's capacities and the nature of the task imposes limits on the subject's information processing capacities. In our terminology, as the number of items increases, the workload associated with the task is increased. The task, however, seems an evident target for analysis in Information Theory terms as it illustrates the manner in which Information Theory may have been used to elucidate the concept of workload.

The analysis of the task begins with its description as a communication task. As seen in Figure 41.1, Pollack has translated the number of tones in a set to a specification in terms of the amount of information per tone in bits.

Figure 41.1 models the subject as a communication channel with a capacity to transmit information that reaches its limits at about 2.5 bits. Similar results were replicated by many investigators, as can be seen in Figure 41.2 taken from Garner's (1962) summary of several studies conducted with different sensory modalities. All investigators seem to find the limit on channel capacity to be at about at 2 to 2.5 bits of information.

Insert Figure 41.2 About Here

2.2.1 Absolute Judgment

A paradigm in which Information Theory appeared to be particularly useful is the Absolute Judgment task. The subject is presented, on each trial, with one of a set of stimuli and is asked to identify that stimulus. Stimuli may vary along a single dimension (e.g., loudness), or along several dimensions (e.g., loudness and pitch). The investigator is generally interested in determining the number of different stimuli that the subject can identify. For example, Pollack (1952) asked listeners to identify tones by assigning a different number to each tone. He varied the frequency of tones, and covered the range from 100 to 8000 Hz in equal logarithmic steps. When a tone was sounded, the listener had to respond with the tone's assigned number.

The results of this experiment are plotted in Figure 41.1. When the set comprised no more than two or three

----

Insert Figure 41.1 About Here

----

stimuli, subjects identified all stimuli correctly. Confusions were rare with four tones, but as the number of tones in the set exceeded four, confusions occurred with increasing frequency. Thus, the subjects were clearly unable to perform a task that is well within their capacity.

human information processing system. In the terms of this discussion, task difficulty is equated in these models with the channel's communication load. The larger the required, relative to available capacity, the greater the communication load. If the human information processing system can be adequately described in terms of classical Information Theory, especially if channel and channel capacity can account for the limitations on performance, the need to invoke mental workload as a concept is obviated. Performance decrements would then be explained in terms of specifiable mechanisms in the language of Information Theory. It is important to examine the degree to which Information Theory has provided satisfactory descriptors of human performance. It would be beyond the scope of this chapter to review in detail the vast literature describing Information Theory in Psychology. The reader is referred to such sources as Coombs, Dawes and Tversky (1970), Fitts and Posner (1967), Keele (1973), and Sheridan and Ferrel (1974). Here we shall briefly discuss studies of the absolute judgments task, choice reaction time, immediate memory, and the control of movements to illustrate the achievements and disappointments resulting from this effort.

(c) Since the information measure is dimensionless it seems likely that Information Theory will provide a general metric by which the processing demands of different tasks can be compared regardless of differences in input modalities, response modes, and experimental conditions.

This enterprise, if successful, would provide both a theoretical framework and a metric for the analysis of those task attributes that lead to invocation of the workload concept. Limitations on the operators are inherent in the framework and a metric derives directly from the Shannon-Weaver (1949) definition of channel capacity. It is instructive, therefore, to examine the efforts made to implement this program and the reasons it has failed to provide a comprehensive solution to the definition and measurement of the limitations on human operators.

2.2 Channel Capacity as a Measure of Limits on the Central Processor

A variety of experimental work designed to test the applicability of Information Theory in the analysis of human performance was undertaken. The common purpose of these studies was to model a variety of tasks in terms of their information transmission characteristics. The studies were designed explicitly to assess the capacity limitations of the

2.1.2 The Putative Advantages of Information Theory

For the Stimulus-Response (S-R) Psychology of the time, the mission was to develop laws to predict responses which organisms make to stimuli. In this context, Information Theory appeared to provide a useful language. It is possible to conceptualize the laws of behavior as descriptors of a communication channel between stimuli and responses. If behavior is organized, consistent, and "lawful," there is mapping from the set of stimuli to the set of responses. The presentation of a stimulus will yield a specific response. In information theory terms we say that information has been transmitted between the stimulus and response. The adoption of information theory and the communication channel metaphor have thus offered to the emerging science of "objective" experimental psychology three main advantages:

(a) Information Theory provides a formal quantitative approach to the modeling of behavior;

(b) Information Theory appears to enable S-R based modeling, with few general, nonspecific assumptions regarding the characteristics of the central processor. Assuming that the central processor operates like a limited communication channel, one could consider workload entirely in terms of the formal properties of stimulus and response.

channel capacity is assumed to have been degraded. The communication engineer equipped with detailed specifications of the communication system can indeed measure channel capacity and can design systems that optimize channel capacity.

The transition to psychology is intuitive but difficult. There were many attempts to borrow the concept of channel capacity in its most precise guise for modeling data on human performance. The appeal of this model as an attempt to develop an "objective" approach to the study of behavior appeared to promise a viable substitute to Stimulus Response models that preserved much of their spirit, as has been noted by G. A. Miller (1956) in his classic paper "The magical number seven, plus or minus two."

> . . . The amount of information is exactly the same concept that we have talked about for years under the name of variance. The equations are different, but if we hold tight to the idea that anything that increases the variance also increases the amount of information we cannot go far astray (pp. 87-97).

2.4.1 Single Bottleneck Models

Single bottleneck views promote an information-flow model which comprises several processing mechanisms, one of which is more constrained than the others. The processing capability of this mechanism then sets the limits for the entire system in coping with task demands. The first and most influential single bottleneck model was proposed by Donald Broadbent in 1958 as a summary statement to his book Perception and Communication. Broadbent proposed the information-flow model summarized in Figure 41.3.

Insert Figure 41.3 About Here

In many ways this is still, quite explicitly, a communication channel model. The central construct was the limited capacity channel (the P system) preceded by a selective filter and a short-term store. This processor leads into a long-term store and a mechanism that selects and controls the system's responses. The filter restricts the entry of information into the processor so that only relevant inputs are fully analyzed. It also affects response selection and transfer of information into long-term memory.

The filter was assumed by Broadbent to operate in a manner that can be described by classical Information Theory.

However, as the communication channel was supposed to model only one component of the system, Broadbent's model was better able to cope with the complexities of the entire information processing system. By postulating that long-term storage and the generation of responses are independent, Broadbent's formulation allows the triggering of responses without intervention by the central processor. Moreover, long-term memory can influence the flow of information through the system. Thus, the metrics provided by information theory could be used to describe the limitations on performance imposed by the filter. Other more cognitive factors could explain the control of the filter. The operation of the filter, in Broadbent's formulation, is influenced by properties of the incoming information as well as by information of long-term store.

Selection by the filter is based, according to Broadbent, on physical features of the input. These are monitored in parallel, but only events that satisfy certain physical criteria are admitted to further processing by the central processor. Broadbent's filter model thus makes strong claims regarding information flow in the information-processing system. Moreover, Information Theory asserts a differential cost to different classes of operations employed in processing

the input. The filter is designed to optimize the information flow given the system's tasks.

Broadbent's bottleneck model is clearly one of workload as much as one of attention. The experimental paradigms within which the model was spawned focused on failures in performance. The filter is, in effect, the source of the subjects' inability to process multiple inputs when the information load is excessive. If workload is viewed as the difference between actual and expected performance given structural definition of a task, then the Workload imposed on a subject by one or many tasks will be determined by the degree to which the information load imposed by the task exceeds the capacity of the filter. (Even though Broadbent's filter is presumably intended to protect a limited capacity processor from excessive information load, the construct regarding which the theory speaks is actually the filter itself rather than the limited capacity processor. In the present context the filter carries the theoretical load.)

The filter can be thought of as an active, early, workload establishing, gate-keeper that limits the information load on the system. The reader will note that there are, in this view, implications for system design. If workload is established by the filter, tasks need to be designed so that they match the properties of the filter. For example, the

nature of the displays associated with a task need to be considered so that requirements of the task are minimal in terms of processing load.

The filter model suggests that additional costs may be incurred when tasks require integration of separate physical attributes in a stimulus array. Consider for example the task described by Treisman and Gelade (1980) in which subjects found it relatively difficult to count the number of blue triangles in an array of blue and white squares and triangles, but relatively easy to count all blue objects or all triangles in the same display. That is, the identification of a conjunction of attributes, requiring parallel processing of the stimuli along several dimensions, imposed a heavier Workload on the system. Issues related to this interesting finding are discussed, at a more formal level, in Chapter 2 by Sperling and Dosher.

Another example of a Broadbentian implication to task design derived from the suggestion that tasks requiring semantic analysis, such as word identification, categorization, or decision making, can not rely on the peripheral filter. The central processing system must be able to deal with these tasks without the presumed protection the central system receives, according to Broadbent, from the filter. Such tasks would benefit from "tagging" all relevant

input units with a common distinguishable physical characteristic since irrelevant units will be rejected from analysis at an early stage.

In summary, Broadbent's model proposed several general principles which made it possible to describe the information flow in the human information processing system, and which serve as anchors for much of the current research in the field of workload assessment. He suggested that the path between stimulus and response might be viewed as three successive stages and described the functional properties of each stage in terms that could be applied to the analysis of the task components. He assigned a cost function to the performance of *mental operations, identifying operations that are more or* less costly to the system. Finally, he emphasized the study of selective attention as an inherent part of the study of workload by suggesting how the efficiency of selection might affect the whole system's workload level.

2.4.2 Welford's Single-Channel Model

We examine another single bottleneck model to illustrate a class of models that focused on the structure of the central processor as a source of the limitations on the system. The difference between Broadbent's model and Welford's (1952, 1959, 1967) model is instructive because both illustrate the

dependence of theories on the experimental paradigms within which they develop. While Broadbent's work was designed primarily to account for such qualitative differences in performance as failures in recall, Welford has been mostly concerned with temporal aspects of performance. The slowing of responses due to the interaction between tasks serves as the primary source of data for Welford's model. Thus, even though Welford, like Broadbent, proposed a single-channel model that was influenced by classic information theory concepts, the two models are quite different.

The key observation which Welford's theory is designed to explain is that the response to a second stimulus is delayed if the subject has not yet responded to a stimulus just presented. The shorter the interval between the first and second stimulus, the longer the response to the second stimulus is delayed. Welford, like Broadbent, postulated a three-stage model of the information flow within the organism, and located the bottleneck of processing in the limited capacity of the central processor. Welford attributed the delay to the limitations on the operation of a central decision mechanism that can process only one task at a time. The time required by the mechanism to process one task was labeled the Psychological Refractory Period (PRP), as a countermatch to the term "refractory phase" used by Teleford

(1931) to describe delays in response of a synapse or a nerve to successive stimulation.

Welford's main thesis was that performance is limited by the operation of a single-channel decision mechanism. This mechanism can deal with data of only one signal, or a group of signals, at a time, so that data from a signal arriving during the reaction time to a previous signal have to wait until the decision mechanism becomes free. The decision mechanism is frequently occupied by feedback from execution of the movements or termination of the response; therefore additional delays may occur even when a signal arrives shortly after the response to a previous signal.

Welford (1967) reviewed data from his work and the work of others to support his arguments and validate the general form of the relationship depicted in Figure 41.4.

Insert Figure 41.4 About Here

These experiments also showed that the delays were not eliminated with training, and that they were localized at the central processor, rather than at a peripheral sensory or motor site. This latter claim is based on the appearance of response delays in the case in which one signal was auditory and the other visual, even when the two responses were made by

different hands, or when no overt response had to be made to the first stimulus (Davis, 1956, 1957, 1959; Fraise, 1957; Slater-Hammel, 1958; Welford, 1952, 1959).

The model also recognized the possibility that a grouping of stimuli may occur if the interval between the stimuli is short. Welford estimated the time required for closure of the gate to be about 80 msec, during which information may enter and grouping can occur. If the rate of an arriving sequence of events is faster than the processing speed of the central mechanism, the response to each event will be proportionately lengthened, perhaps enough that some stimuli will not be processed. A temporary buffer was assumed to store at most two events, thereby lengthening the period of graceful degradation.

Insert Figure 41.5 About Here

Figure 41.5 illustrates the manner in which Welford's model was used to interpret the results of an experiment by Conrad (1954), requiring subjects to respond by pressing a key or turning a knob, each time one of a number of rotating pointers coincided with one of several irregularly spaced marks on the edge of dials. The number of dials (2, 3, or 4) and the speed of rotation of the pointers were used to vary

the information on the display. When the level on each of these variables was increased the portion of events omitted was increased and responses were delayed.

The reaction time data were fitted with a model that assumed a constant reaction time for a given number of dials. This number was converted to bits of information and related to the Hick-Hyman formula for the determination of an appropriate reaction time value (see Chapter 30 by Keele). As can be seen the fit of the equation to the actual data is quite convincing.

It is instructive to compare the notions of central limitation and workload that emerge from Welford's single-channel operation with those of Broadbent discussed earlier. Both models rely heavily on the Information Theory paradigm, both propose three main stages in the flow of information, and both locate the bottleneck in the operation of the central mechanism. However, they differ substantially in their assumptions regarding the operational rules governing this mechanism. Broadbent emphasized a limit in terms of the total amount of work per unit time. His model allows parallel processing as long as the joint demands of tasks do not exceed the limit of the central processor. Additional tasks can enter the central processor at any time if the currently processed task does not exhaust the capacity of the processor.

The limits discussed by Welford are more structural, in the sense that the mechanism is completely inaccessible to serve in the performance of other tasks when it is occupied by one task. Parallel processing depends on the time of arrival of tasks and the ability to group them. When grouping is possible the tasks in effect become a single task that is likely to occupy the central processor for a longer duration.

The implication of this type of limitation is that once a task or a task element enters the decision mechanism its processing proceeds uninterrupted. Partial processing or errors due to interactive influence of elements are assumed to be rare. The decrements in performance that are manifestations of workload appear as prolonged response times or as omissions of the response. In contrast, partial processing and interaction errors are consistent with the model of performance limitations proposed by Broadbent. This difference between the models may result from different experimental paradigms and the principal dependent measures from which each of these models derived.

These issues were not addressed by Welford and, in retrospect, were of less importance within the framework of the experimental tasks Welford used. All messages were assumed to be processed in the order of arrival and to depend on the processor's ability to complete its earlier commitment.

Problems of selection and load did not require a direct consideration of task components and interaction between tasks.

Even at this early stage of the analysis of limitations on system performance we encounter an interaction between measures of performance and the model. Reductions in "quality of performance" can be assessed in a variety of ways, therefore that an operator performing a task may appear to be either fulfilling or failing to fulfill performance expectations depending on the measure of performance selected by the observer. Much of the controversy in this field may be traced in part to such different perspectives on performance. A celebrated controversy that confronted "early" selection with "late" selection in attention, with Broadbent favoring early selection and Welford's model implying late selection, may be due in large part to different perspectives on performance from the different vantage points of dichotic listening and reaction time studies. We will carry this cautionary note into further discussions of workload in the remainder of this chapter. Performance is in the eye of the beholder and workload measures are limited by the perspective of their design. Performance quality varies with the type of performance being studied and there is a corresponding limitation to performance-related measures of workload.

2.4.3 The Breakdown of the Single Channel

Both single-channel models were provocative enough to stimulate extensive experimental work, but both lacked the power to account for the growing body of data. The ultimate test for any model of information flow within the organism is its ability to relate the functional properties of the model to the characteristics of human tasks thereby improving the prediction of behavior.

Welford's single-channel operation could not explain, for example, the finding that varying the difficulty associated with response to the second stimulus affects the pattern of responses to both stimuli (e.g., Karlin & Kastenbaum, 1968; Keele, 1973). If the second stimulus is not examined before analysis of the first stimulus is completed, how can its difficulty modify the interval between the first and second response?

Another criticism of Welford's model was suggested by Kahneman (1973). He argued that instead of examining the change in speed with which subjects respond to the second stimulus as a function of the interval between the first and second stimuli (ISI), one should examine the interval between the first and second responses (IRI). When this measure is employed clear evidence emerges for processing in parallel of

the two stimuli. That is, the second stimulus does not wait until analysis of and response to the first one are completed.

Similarly, experimental tests of predictions derived from Broadbent's model yielded evidence that is inconsistent with the early selection hypothesis, and with the idea that the filter serves to protect the central mechanism from possible "overload" of high-level semantic processing demands. These assumptions were challenged by two main groups of findings, the first of which addressed the assumed costs of semantic processes. It showed that subjects were able to follow, with little impairment in intelligibility, verbal messages presented at twice the normal rate (Fairbank, Guttman, & Miron, 1957). Also, if the information content of a passage was doubled by using low-level approximations to English, shadowing performance did not decrease to 50% (Treisman, 1965). These findings conflict with Broadbent's interpretation of the inability to follow the verbal content of the message on the irrelevant ear in a dichotic listening task as indicative of the high demands imposed by verbal processing on the central processor, precluding simultaneous listening and processing of material from the two ears.

Another series of studies examined the claim that filtering by physical cues discards the irrelevant information from further analysis. It was found in "shadowing"

experiments that if the text presented on the relevant ear was switched to the irrelevant ear, subjects followed the text and abandoned the information presented on the relevant ear (Treisman, 1960). Other experiments have shown that subjects could detect significant words on the irrelevant ear (Moray, 1959, 1967), or monitor for target words on the two ears. All the above could not have happened if Broadbent were correct in assuming that information on the irrelevant ear does not reach the level of semantic analysis.

These findings led several researchers to propose a late rather than an early selection bottleneck in the flow of information (e.g., Deutsch & Deutsch, 1963; Keele, 1973; Norman, 1968). The main contention of this alternative was that encoding and intake of information are not as demanding and can be handled in parallel. The system becomes single channeled and limited once the process of evaluating information and selecting an appropriate response has begun. Data in support of this argument were obtained in experiments in which probes are inserted during an interval (e.g., Posner & Boies, 1971). The response time to a probe stimulus presented during the processing of a primary stimulus turns out to be unaffected during the first 300 msec after the presentation of the primary task, but the response is delayed considerably if the probe is presented later in the interval.

limitations and consider their appearance to reflect a decline in the level of a hypothetical entity, the "resource." This usage of the resource concept is not much different from the use one makes of "arousal" which, after all, is a hypothetical construct drafted to account for the evident differences in state that we display as we move on a continuum from slumber to activity. Something clearly changes when one wakens from sleep. The change is most properly described by reference to a host of specific changes in neural circuitry and in the levels of a variety of neurochemical substances. Yet, for a variety of purposes one may ignore this complexity and describe the organism as moving along a dimension of arousal. This is a useful theoretical device as long as one does not infer from its success that inside the body is a specific system that embodies, or secretes, an entity corresponding to arousal.

A discussion of performance limitations organized in terms of resources is not a structural model of the information processing system. Rather, it is an assertion that it is possible to treat the system, for the purpose of describing its limitations, as if it depends on the availability of some hypothetical resource. In a sense, we are implying by use of this concept that it is not quite necessary to know <u>why</u> an operator fails to perform. As long

Activation Theory (Lindsley, 1951). This psychophysiological model assumes a structure of analyzers and processors that is fixed in its capacity. Processing is possible to the degree that any given processor is "activated." The activation, in Lindsley's model, depends largely on the influence of the reticular activating system that emerged from the studies of Magoun (1949) and his collaborators immediately following World War II. There is a fairly direct route from the "physiological" activation that underlies the ability of the system to process information to Kahneman's resources.

It is critical to understand that neither activation nor resources are directly observable entities. In both cases these are hypothetical constructs introduced to organize observations on performance. This point is sometimes neglected in discussions of "resource" models. The literature tends to slide into a reification of resources in a manner that suggests that the concept refers to actual observable entities; it is better to admit that the "resource" is a concept that does not relate to a specific structural model of the system. In adopting this concept one asserts that organisms behave as if they are resource-dependent systems in the sense that there are clear limits on their capacity to perform. As it is rather difficult to pin the limitations on any specific mechanism, it may be useful to pool the

such as "resources," is consistent with the view of workload presented in the introduction to this chapter.

If workload is a hypothetical construct designed to account for the variations in the function of the operator-task loop, then the test of the utility of a construct is determined by the descriptive economy it allows rather than in its fidelity as a structural model of the system under consideration. As the next section will show, energy and resource models do provide rather attractive descriptors of the system's functions, though here again the system will require substantial elaboration of these models.

2.6 Energy Constraints on Processing Capabilities

2.6.1 The Single Resource Model

The most comprehensive attempt to employ the energy metaphor in the analysis of workload, and of attention, can be found in Kahneman's book Attention and Effort (1973), which attributes a system's failure to perform to a shortage in the supply of what he calls processing resources. "Resources" is a label applied to a single undifferentiated pool of energizing forces necessary for task performance. In its derivation, the concept of resources is closely related to the concept of arousal, which has an important role in some theories of attention. Consider for example Lindsley's

technique is that it provides a methodology for operationalizing an analysis of a task's structure in terms of task variables. This view of the system is important in some of the common approaches to the measurement of workload. The notion that one must consider the system as having a structure underlies the approach to workload that requires a formal task analysis in terms of its specific "structural" demands as a precondition to analysis of task workload. This contribution is not vitiated by the fact that it is obvious that the human information processing system shows extensive concurrency and interaction among its elements.

The apparent complex architecture revealed by contemporary analyses of the information processing system is costly. The complexity of the structure, multiplied by allowing concurrency and parallelism into the system, has made it difficult to provide simple tools for workload measurement. It is possible, however, to ignore the structural complexity of a system when trying to assess its limitations if the critical limitation can be ascribed to a uniform source viewed independently of the system structure. Such a solution is provided by adopting yet another metaphor, the "energy" metaphor. In analyses of workload, the energy metaphor has played a dominant role, and it is reviewed in the following section. The energy metaphor, as well as related metaphors

The general structure of information flow and the logic underlying the additive factors methodology were generalized to other tasks and to different problem areas such as workload assessment and dual task performance (e.g., Logan, 1978, 1979; Whitaker, 1979; Wickens, 1980) as well as task analysis (Mane, Coles, Wickens, & Donchin, 1983). The paradigm and its assumptions and findings played an important role in the formulation of an elaborate model of the central processor whose limitations are, or whose principal interest is in, the analysis of Workload.

As Treisman's elaborated filter demands a multifaceted view of workload, so does a view of the system that locates the effects of different task attributes in different processing stages. Neither of these models is consistent with a notion that workload is due to a specific and unique deficit in the system. A task may vary in ways that limit performance because one or several stages of processing are affected. The effect of variations on the structure of a task on the demands it imposes may be additive for some variables and interactive for other variables. It may also happen that task demands will not be affected by a manipulation of task structure. Some variations in task demands may be totally irrelevant to performance of one task, yet be crucial to performance of another. One of the major advantages of the additive factors

Table 41.1 Summary of experimental variables that were found to have additive and interactive effects on choice reaction time. Based upon the rational of the additive factor methodology, those variables that interact influence to the same stage, while those that are additive affect separate stages. (Adapted from Sanders, 1980)

Additve effects:
Signal quality x S-R compatibility: Sternberg (1969), Shwartz et al. (1977), Frowein and Sanders (1978), Sanders (1979).
Signal contrast x S-R compatibility: Shwartz et al. (1977), Sanders (1977).
Signal contrast x signal discriminability: Pachella and Fisher (1969), Shwartz et al. (1977).
Signal contrast x signal quality: Frowein (note 2), Sanders and Akerboom (Table 3).
Signal contrast x word frequency: Becker and Killion (1977).
Signal quality x word frequency: Stanners et al. (1975).
Signal discriminability x S-R compatibility: Fisher and Pachella (1969), Shwartz et al. (1977).
S-R compatibility x Instructed muscle tension: Sanders (1979).
Signal quality x Instructed muscle tension: Sanders (1979).
S-R compatibility x Response specificity: Sanders (1970).

Interactive effect:
Signal quality x Movement frequency x Movement Predictability: Wertheim (1979).
Stimulus contrast x S-R compatibility: Stanovich and Pachella (1977).
Stimulus contrast x Meaningfulness: Miller and Pachella (1976).

Priming x Word frequency: Becker and Killion (1977).
Priming x Signal quality: Meyer et al. (1975).
Priming x Signal contrast: Becker and Killion (1977).

consequent changes in reaction time to establish the internal stages, unobservable by themselves. It was Sternberg (1969), however, who incorporated this logic in the general framework of information processing and developed a formal approach to the extraction of stages. Sternberg reasoned that variables having an additive effect on response time must operate at different stages, while variables that interact influence at least one common stage. For example, signal quality and stimulus-response compatibility were shown to have additive effects on reaction time in a variety of situations (e.g., Sanders, 1979; Shwartz, Pomerantz, & Egeth, 1977; Sternberg, 1969). In contrast, stimulus response compatibility was found to interact with time uncertainty (Broadbent & Gregory, 1965; Sternberg, 1965). Signal quality and compatibility were interpreted to affect the stages of encoding and response choice, respectively, while compatibility and uncertainty were both associated with the response selection stage. Table 41.1 summarizes variables found to have additive or interactive effects on response time.

The term "processing stage" thus refers to an aggregate of processing structures or computational processes that represent a common mental operation. Thus the preprocessing of a stimulus, feature extraction, response choice, response programming, and motor adjustment are all examples of "stages" (Sanders, 1980, 1983). The number of stages is not fixed and depends on the requirements of the specific task. For example, assume that the subject is required to determine if two letters in a pair are "the same." If the decision is based on the character and on the font in which the pair is presented (so that Aa and AB may be considered different while AA and aa call for the response "same") then there is a need for a stage in which the physical features are compared and one in which the letters are identified. In contrast, if the font alone is the criterion for responding (for example, AB and ab are the "same" and Aa and Bb are different) then only physical features need to be identified (Posner, 1978).

The main experimental tool in the analysis of the string of stages for a given task is the prolongation of response time as a result of manipulating task variables. In the majority of studies, such an analysis is based upon the additive factors paradigm proposed by Sternberg (1969). Donders suggested in 1868 (see reprint in Donders, 1969) that it is possible to manipulate task variables by measuring

2.5 The Architecture of the Central Processor

As it became evident that the limitations on performance require an examination of the structure of the central processing system, it became necessary to develop a methodology that could be applied to the analysis of the system. A natural extension of the models reviewed is the identification of multiple components in the system each of which is describable in simple terms and with simple rules for governing the interactions between them. The components tended to be viewed as functional entities each playing an identifiable role in the informational transactions of the system. A simple rule for the interactions between components is that they are strung in sequence, each feeding its successors with information. Such a view of the system leads naturally to the concept of a "stage," with each component serving as a stage in a sequence of processing. According to this model, the operation of a successive stage does not begin before preceding stages have completed operations (Sanders, 1980; Sternberg, 1969). Other approaches are possible. Thus McClelland (1979) proposed a Cascade model and Eriksen and Schultz (1979) a continuous flow model according to which all stages of processing operate continuously, passing information from one stage to the next as the information becomes available.

when it refers to an ensemble of processing entities that communicate with each other under complex control schemes. Thus the concept of what constitutes a channel has been considerably changed.

Regardless of the remaining viability of the channel concept, the decades following the predominance of models derived from communication theory were characterized by a shift to an examination of the internal structure of the processor with a specific interest in its architecture. We proceed to review this phase. It will be seen that a dominant concept in this paradigm was "stages of processing." The relevance of this literature to the analysis of workload may not be obvious. The concern with architecture reduced the prominence of the processes that account for limitations of processing. However, as it became clearer that the architecture of the system is complex it became evident that limitations in processing can appear in a variety of nodes in the system, and the cost of different mental operations emerged as a central issues in the attempt to link the limitations of the central processor to the components of task demands. The following section reviews studies of the architecture of the processing system (e.g., Sanders, 1980; Sternberg, 1969).

implications for the processing operations carried out between stimulus and response. Tests of relevance could now be performed at the input, the dimensions (analyzers), the memory and response, or the item levels. The model represents a further step in the decomposition and elaboration of the original global notion of information. An implication of this approach to the study of workload is that the assessment of the workload imposed by tasks must now be detailed further in terms of the components of processing involved. This approach is particularly significant as it is related to the degree to which parallel rather than sequential processing is the dominant mode. In a system comprising many mechanisms and processes, elements of a task are likely to simultaneously occupy different processing mechanisms, and to be at various processing states. In such a system the original notion of a single flow, sequential streaming of information loses much clarity and power. This view of the limitations of the information processing system leads naturally to a view of workload as a multifaceted concept as there are various reasons for a decrement in performance.

It seems evident that even though Treisman's variant of the filter model is a "single channel" model, it actually permits a rather elaborate definition of the channel concept. In fact, the very notion of a channel loses much of its value

The current versions of this approach tend to resemble the version of the filter model proposed by Treisman (1964, 1969). Triesman's model differed from Broadbent's in two major aspects: (a) the filter was assumed to attenuate rather than completely block the information arriving in the unattended channel; and (b) the filter was given a strategic flexibility, in the sense that it could be deployed at various locations along the information processing path, to increase the cost effectiveness of the selective process. In Treisman's words:

> . . . Four types of attention strategy are distinguished: The first restricts the number of inputs analyzed; the second restricts the dimensions analyzed; the third the items (defined by sets of critical features) for which subject looks or listens; and the fourth selects which results of perceptual analysis will control behavior and be stored in memory. (Treisman, 1969, p. 282)

Note that for Treisman the filter is an "attention strategy" rather than a fixed structure, implying that filters can be activated at different locations depending on the conditions at which a task is employed, and the nature of the competing environment. This later argument has testable

system, its rejection as a general model should not imply a rejection of the possibility that there are such "early" selection processes. The observations on which these models were based cannot be ignored. Thus, for example, the dramatic loss of information arriving in the unattended ear in shadowing tasks is real enough, as are the difficulties encountered when attending to and integrating simultaneous sources of information. Furthermore, rather dramatic direct evidence for the operation of an early selection has been provided by Hillyard and his colleagues (Hillyard, 1984). These investigators recorded event-related brain potentials (ERP) from human subjects while these subjects were engaged in Broadbentian multiple channel monitoring. For a definition of ERPs and a review of the manner in which they are recorded and interpreted see Section 3.9. The ERPs recorded from the attended channel were distinctly different from those recorded from the unattended channel. Moreover, these differences appeared as early as 100 msec after the eliciting stimulus. Naatanen (1983) presented a theory of selective attention that integrates these ERP data with the literature reviewed in this section. Other attempts to defend the validity of an early selection process can be found in Broadbent (1982) and in Kahneman and Treisman (1983).

It is as if the interference was most powerful beyond the point at which the main processing was completed, and the subject was involved in "thinking" or deciding about the proper response.

The inadequacies of these bottleneck models were also implied by various demonstrations that the recall of information (which implies that the information had been processed and represented in memory) was more demanding than the intake of multiple inputs (e.g., Martin, 1970; Trumbo & Milone, 1971). Thus, the notion of a bottleneck early in the processing stream gave way to models of processing that assumed a more central ("late" in the jargon of the times) locus for the limitations. We see here the same process that affected the utility of information theory models. The information processing system, upon close examination, is active and dynamic and any attempt to model it with a peripheral, data driven model is unlikely to capture its full capabilities, even if one allows the peripheral filter to be controlled by central factors. Adequate models of attention, and by implications of workload, require consideration of the internal structure and the operating modes of the information processing system.

Even though an early bottleneck model fails to provide a full account of the operation of the information processing

as we can quantify the deviation between expectation and actual performance we can treat all failures alike. This caveat is in place whether one adopts the uniform-resource model presented by Kahneman (1970) or some version of the multiple resource models we will review in Section 2.6.2.

Kahneman viewed the amount of resources available at any time as limited, but the limit varied with the level of arousal, according to the classic inverted U function relating effectiveness of performance to arousal. Changes in the level of arousal and consequent changes in capacity are assumed to be controlled by feedback from the execution of ongoing activities; a rise in these activities causes an increase in the level of arousal, effort, and attention. The general structure of the model is depicted in Figure 41.6.

Insert Figure 41.6 About Here

An important construct in Kahneman's model is the mechanism responsible for the allocation policy. This mechanism directs and supervises the allocation of resources and is influenced by enduring dispositions, momentary intentions, and the feedback from ongoing activities.

In a structural model, failures to perform occur when a mechanism is required to carry out incompatible operations.

For example, one's eyes cannot monitor simultaneously two screens placed on opposite sides of a room. This limitation on performance is attributed to the specific structure of the human visual system. In an energy-oriented model, such as Kahneman's capacity model, decrements in performance are due to demands of two concurrent activities exceeding the available capacity. A structural model therefore implies that the interference between two tasks is specific. In a capacity approach it is nonspecific, and depends only on the total demands of the two tasks.

A concept of a central limited source of processing energy provides a convenient solution to the conflicting data on the location of the bottleneck in the information flow within the processing system. Moreover, the concept is consistent with the conclusion that there exists a fairly general source of limitations on the central processor. Such a general source of limitations appears necessary since one often observes interference between tasks that have very little in common. Why, for example, should there be any interference between the ability to recall and the difficulty of a simultaneous manual tracking task (Johnson, Schulman, Greenberg, & Martin, 1970)? Or, why should the ability to attend to auditory messages be affected by the need to monitor a visual display for a critical letter (Kahneman, Beatty, &

Pollack, 1967)? This interference may be explained by invoking the concept of a general energy source of a fixed capacity made available to one or another task but not to both. However, this source of energy is not observable. No direct observations on the level of available resources have been proposed either by Kahneman or by others.

This aspect of the resource concept tended to lead its proponents to rely on observations on physiological systems in an attempt to monitor the level of available resources, or the level of demand on the resources. Such attempts were made by several researchers prior to Kahneman, (e.g., Berlyne, 1951, 1960, 1970; Easterbrook, 1959; Wachtel, 1967). However, none offered as detailed a model. Kahneman's measure of effort derived mostly from his own data on the effects of mental effort on pupil dilation. Several experiments conducted by Kahneman and his colleagues (Kahneman et al., 1967, 1968, 1969, 1971) showed changes in pupillary response that paralleled variations in task demands. For example, when subjects had to memorize and transform a list of digits, the pupil was largest when the demands on memory were highest. Pupil dilation was greater when the transformation the subjects were to perform on the digits was more difficult (Kahneman, Peavler, & Onuska, 1968).

These data demonstrate the sensitivity of the pupillary response to processing effort, and serve to demonstrate the competition for "energy" between tasks with little structural similarity. Additional support for the relationship between the pupillary response and processing effort has come from recent studies by Beatty and his associates, summarized in Beatty (1982), who continued and elaborated Kahneman's research.

With a notion of flexible limits on processing capacity that change with feedback from behavior (which in turn influences the level of arousal), an independent physiological measure of processing demands has a crucial role. Performance measures by themselves are poor indicators of resource limitations, because performance is both the result of the limitation, and a trigger of a change in the limit via recruitment of additional resources. Only an independent measure can break this vicious circle. The main test of the model is the validity of the claims that all tasks, regardless of structure, compete with each other, and that an increase in the difficulty (demands) of one task will be reflected in the ability to perform another one simultaneously. By and large, it seems fair to say that the single-resource model did not fare well under experimental analysis.

Data from a variety of experimental conditions have shown that performance on some tasks interferes with one type of task, but not with another, while the third group is equally affected when paired with members of the first two groups. For example, mental arithmetic is little impaired when jointly performed with a pursuit tracking task, but severely degraded if paired with a choice reaction task to visually presented digits. In contrast, the choice reaction task is equally degraded when paired with mental arithmetic or pursuit tracking (see Navon & Gopher, 1979; Ogden, Levine, & Eisner, 1979; Wickens, 1980, 1983, for this and additional examples). Other studies have shown that some manipulations of task variables within the same pair of concurrently performed tasks affect both tasks, while others degrade performance on one task only, and cannot be compensated by shifting resources from the performance of the shared task (e.g., Gopher & Navon, 1982; Gopher, Brickner & Navon, 1982; Wickens & Kessel, 1981). These results are inconsistent with the notion of a single undifferentiated pool of processing energy. Rather, they suggest the existence of several more specific sources of interference and competition. Note that we are again driven by the data to the conclusion we drew from the analysis of single bottleneck models. Again, the concept of a single

general cause of performance limitations attributed to Workload must be rejected.

2.6.2 Multiple Resource Models

The energy, or the resources, metaphor is an attractive approach to the analysis of workload. It has an affinity to the origins of the concept in the ergonomists' physical workload and it is an intuitively appealing way to describe what is lacking when performance falters for "no apparent reason." The evident weakness of the single resource model leads therefore to the development of multiple resource models according to which the human system is best modeled as possessing a number of processing mechanisms, each requiring its own supply of "resources." The capacity of each of the structures, that depended on the level of arousal and its own specific dependence on this level, can be deployed at any moment among a number of tasks. Thus, there is continuing competition for resources between tasks that overlap in resource needs, (Gopher & Sanders, 1984; Norman & Bobrow, 1975; Navon & Gopher, 1979; Sanders, 1983; Wickens, 1980, 1983).

Within this framework the key questions are: what is the nature of resources? How are they related to one another? How can the demand composition of tasks be expressed in terms

of the underlying resources? Norman and Bobrow (1975), the first to employ the term "resources," were not very specific. They used a general analogy to a computer system, and resources were defined in general terms with reference to all processing facilities. Specific examples were "...Such things as processing effort, the various forms of memory capacity, and communication channels..." (p. 45). Note that in Norman and Bobrow's writing the term "resource" lacks the energy-like flavor of Kahneman's description of the system. Rather, the system appears to be described by Norman and Bobrow on a structural level. Yet, even this seemingly structural approach is nonspecific in modeling the nature of the deficits. Once performance limitations are not attributable to any deficiencies in the supply of data, they are attributed to an inadequacy of resources.

Navon and Gopher (1979, 1980) who also elaborated the multiple resource framework, were not more specific in their initial theoretical analysis. They employed an economic metaphor and drew an analogy between the problems facing a person performing one or two tasks and a manufacturer of one or more products who has to optimize the use of his resources (such as labor, equipment, raw materials, etc.). In subsequent experimental work (Gopher & Brickner, 1980; Gopher et al., 1982) they conclude that it is reasonable to assume

the existence of at least two relatively independent types of resources; one is related to perceptual and computational processes, and the other is linked with selection and generation of motor activity.

The nature of resources is a central issue of concern in the work of Wickens (1980, 1981, 1983) and his colleagues (Wickens et al., 1981, 1982, 1983). Wickens (1980) proposes three plausible candidates for the structural composition of resource reservoirs: stages of processing, cerebral hemispheres, and modalities of processing (both encoding and response). Classification by processing stages follows the architecture of the processing system as it emerges from experiments based upon the additive factors methodology (Sanders, 1979, 1980; Sternberg 1969). Hemispheres of processing are suggested from theoretical analysis and experimental works that view the cerebral hemispheres acting partially as separate resource reservoirs by virtue of their functional and spatial separation (e.g., Friedman & Polson, 1981; Friedman, Polson, Defoe, & Goskill, 1982; Kinsbourne, 1975; Kinsbourne & Hicks, 1978). The direct relevance of this literature to task analysis bears more careful examination (Donchin, McCarthy, & Kutas, 1977).

Justification of modalities of processing and response as a classification criteria is provided by those studies which

compared auditory and visual modes of presentation and verbal versus manual modes of response in dual task paradigms (e.g., Gopher, Brickner, & Navon, 1982; McLeod, 1972, 1978; Wickens & Kessel, 1981; Wickens & Sandry, 1982). Wickens (1981, 1983) proposes a three-dimensional descriptive framework that can serve to organize these categories (Fig. 41.7) and also to include a distinction between types of representation codes, verbal and spatial.

Insert Figure 41.7 About Here

The framework proposed by Wickens summarizes factors that demonstrably influence the pattern of interference between two concurrently performed tasks, and are linked with a competition for access to a central mechanism. However, it says very little on the way energetical and structural elements are related to each other. This relationship between energy and structure is the major concern in the cognitive-energetical stage model proposed by Sanders (1983) and Gopher and Sanders (1984) (Fig. 41.8).

Insert Figure 41.8 About Here

The cognitive-energetical stage model is an attempt to integrate energetic concepts with a structural description of the system. Its development derived from data on the effects of stressors on performance in choice reaction tasks (e.g., Frowein, 1981; Sanders, Wijmen, & V. Arkel, 1982). This research showed that the effects on response time of stressors such as sleep loss, time on task, or psychoactive drugs are selective. The stressors affect specific mechanisms but have no general effect on performance. For example, amphetamines were found to interact only with those variables associated with motor activity, while barbiturates interacted only with variables that influence processing activity as related to feature extraction.

These findings were interpreted within the framework of the neurophysiological model of attention control proposed by Pribram and McGuinness (1975) who identify three main energetical generators of processing activity: Arousal, Activation, and Effort. Integrated within the proposed structure of processing stages, "arousal" is linked to input encoding activity. "Activation" is argued to energize output processes and "effort" is given the role of activating the central decision making and choice mechanisms. In addition, effort is assumed to secure an optimal level of operation of the other two energy sources and to act as a coordinator

between their activities. Whether the effort level by itself is in an optimal state depends on evaluation of the state of adaptation of the organism to environmental demands. In other words, effort invested depends on motivation and on an assessment of the situation. There is a resemblance between the role of the evaluating mechanism in the cognitive energetical stage model, and those given to the allocation policy construct in Kahneman's (1973) single capacity model, although in the present model the evaluation mechanism has an independent link with each of the three energy mechanisms, and hence may have a separate influence on different aspects of performance.

With an increase in overall complexity, and with less simplicity than was typified in the original single-channel and central-capacity models, an integration of these complementary frameworks may possess the degrees of freedom necessary to cover the processes limiting the work of the central apparatus. Which of the identified structures and energy sources will be a productive descriptor of an underlying processor is still unknown. The two frameworks were developed as a post-hoc interpretation of different bodies of data. Preliminary research has supported a distinction between perceptual and motor resources (Gopher & Navon, 1980; Gopher, Navon, & Brickner, 1982; Wickens &

Kessel, 1982). Studies by Wickens and his colleagues, and experiments from other laboratories report encouraging results concerning the separation of tasks along modalities of processing and types of representation codes (Baddely & Liberman, 1980; Brooks, 1967; Moscovitz & Klein, 1980; Reisberg et al., 1984; Vidulich & Wickens, 1981). Initial data from measurement of the event-related electrical activity of the brain during task performance (Donchin, Kramer, & Wickens, 1983; Israel, 1980; Wickens, Kramer, & Donchin, 1983) and experiments on the application of stressors mentioned earlier in this section support arguments for the existence of several sources of energetical activity. However, the general state of this knowledge is still preliminary. Even so, the formal properties of models akin to multiple resource models are being clarified (see Chapter 2) by Sperling and Dosher.

Our survey of approaches to the limitations on the information processing system concludes by considering an important aspect of performance neglected by most of the models reviewed, the influence of practice on performance. It is well known that practice is the single most powerful factor improving the ability of the system to perform a task. Thus, nothing is as likely to reduce associated workload as is practice. Yet few of the models of attention and performance incorporated the effects of practice on workload and its

interaction with other factors that contribute to workload. The framework in which this factor is beginning to receive its due respect is in the analysis of the distinction between "automatic" and "controlled" processing (Schneider & Shiffrin, 1977).

2.7 Controlled and Automatic Processes

Reduction of workload with continued practice has been attributed to development of automatic links between stimulus and response that can be operated with minimal interaction from the central processor. Automatic processing is defined as a fast parallel process not limited by short-term memory, requiring minimal processing effort, amenable to little direct control by the subject, and requiring an extensive and consistent training to develop. Walking, speech production, and driving after years of experience are examples of highly automated behaviors.

Controlled processing is relatively slow, is mentally demanding, requires considerable involvement of short-term memory, exhibits a large degree of voluntary control by the subject, and requires little or no training to develop. In recent years there has been considerable interest in the comparative study of these two processes, accompanied by a detailed analysis of the ways in which automaticity is

developed (e.g., Laberge, 1975, 1981; Norman & Shallice, 1981; Schiffrin & Domais, 1981; Schneider, Dumais & Shiffrin, 1983; Schneider & Fisk, 1983).

The main requirement for automaticity is the existence of consistent mapping (CM) between stimulus and response. For example, in a task requiring the identification of target letters in an array of distracting letters on a visual display, those letters employed as targets should never become distractors, and vice versa. Under such conditions, a high level of parallel processing is developed, and reaction time is reduced and appears to be independent of the number of distracting letters on the display (e.g., Schneider & Shiffrin, 1977). In contrast, if the mapping is frequently changed around, that is, targets become distractors and distractors targets (VM), response times are slow, the slope of the training curve is shallow, and there is a linear increment in the time to identify a letter with an increase in the number of distractors on the display (see Fig. 41.9). The

----

Insert Fig. 41.9 About Here

----

same phenomena have been observed when a task calls for a distinction between words, or identification of noun membership in semantic categories (Fisk & Schneider, 1983).

to document the structure of a task. If we assume a normative operator, the designer of a system may be affected in the choice of assignments given to the system's operators by the number of activities called for according to the time-line analysis. The assessment of the degree to which a task may or may not be more difficult, as judged by the time-line record, depends on our assumptions regarding the capacity of the operators. In other words, time line analysis can not be considered a procedure for the assessment of workload except for those restricted circumstances in which the task will be presented solely to individuals who fit certain selection criteria. Moreover, the assumption is also made that there will be no interaction between the task and the operator that changes the character of our statement.

Insert Figure 41.10 About Here

From the analysis presented in Section 2 it should be evident that it is unlikely that a global standardized measure will capture the complexity of the human information processing system. This is true at least as far as the limitations on the system are concerned. Much as the mystique of the super-human experimental pilot is appealing, the reality is that all humans are equipped with limited capacity

workload was assessed, then the measurement of workload for this class is complete. The problem is that the trade-off between design and selection can not be avoided. Designers can develop systems that can be used, at a minimal workload, by most operators, or systems can be designed for effective use by the select few. In the first case the design may be costly, but the workload is small. In the second case the design may be simpler but the cost of selection and training may prove excessive.

The assumption of a standard operator is typical of a normative approach to the measurement of workload. In a normative approach, one assumes that the operators satisfy some criteria for capacity, motivation, rationality, and other related attributes. The task is analyzed for its challenges to the standard operator and conclusions are drawn regarding the workload in the loop. This approach underlies what, in industry, is probably a prevalent method for assessing workload. We refer to time line analysis, an inherently open-loop technique, which focuses on the structural description of tasks. Its practitioners have acquired the skill to examine tasks rather minutely and to identify the components of the task. The specific actions that must be taken at any instance along the epoch of task performance are ascertained. The product is a chart, illustrated in Figure 41.10, which serves

therefore cannot measure workload. Section 1.1 mentioned leaping across the Grand Canyon as a clearly difficult task with respect to which workload will not be measured, because our knowledge of the capacity of the operator precludes this particular loop from consideration.

It is critical to distinguish between measurements of subject capacity or analyses of a task's structure and the measurements of workload. The critical distinction lies in the degree to which workload measures are made in the context of, and refer to, the interacting combination of subject and task. Strictly speaking workload must be defined anew for each subject, and for each set of prevailing circumstances. We relax this rule on the assumption that an individual can be assumed to remain (subject to the effects of training and practice) essentially stationary. Thus, generalizations can be made across occasions.

3.2.1 Normative and Descriptive Approaches

Assuming for design purposes that human operators are a homogeneous class, a properly selected group can be considered to satisfy certain capacity criteria and workload assessment can be accomplished for these operators even if only a few representatives are assessed. Moreover, if the task is then assigned only to individuals within the class for which

catalog of available measurement techniques. Here we discuss only a few techniques that illustrate classes of problems needing attention in the design of workload measures.

3.2 Workload as a Property of the Operator/Task Loop

The most fundamental assertion regarding the measurement of workload is that workload is an attribute of the loop between an operator and a task. Workload is a hypothetical construct intended to capture limitations on the operator's information processing apparatus as these are viewed from the perspective of some assigned task. The critical implication of this assertion is that it is not particularly meaningful to measure workload in an open loop. One can not specify workload associated with a task without reference to the operator, and one can not relate workload to an operator without this operator's being in the matrix of the task/operator loop.

Relevant operator and task characteristics having an important effect on workload characterize a closed loop. The operators' capacities and skills can be assessed and the structure of the task designated a priori or ascertained by means of a task analysis. When the information about the task or the operator suggests that the task/operator loop will accomplish nothing we do not bother to close the loop and

proposed approach is undoubtedly less concise and parsimonious than one would like, but is a simpler one possible, given the richness and complexity of human mental activity?

3. TECHNIQUES FOR THE MEASUREMENT OF WORKLOAD

3.1 Criteria for the Evaluation of Workload Measures

This section discusses classes of measures that have been proposed for, and are employed in, the measurement of workload. All the procedures reviewed have been proposed explicitly for the measurement of workload regardless of theoretical underpinnings of the measurement. Measures of workload have been developed, for the most part, in a pragmatic context. This explains, in part, the lack of theoretical consistency in the analysis of this concept. The approach has often been intuitive and measures have rarely been evaluated for their validity and reliability. Indeed face validity (a term that lends dignity to an unquestioning reliance on one's intuitions) is often considered paramount in evaluating workload measurement. Thus not all procedures that purport to adequately measure workload will satisfy the criteria derived from the analysis in this chapter. It is our intent to review measurement techniques within the framework of our own interpretation of workload. The reader is referred to O'Donnell and Eggemeier, Chapter 42, for a fairly complete

In conclusion, what are the general guidelines of the theoretical analysis carried out in this chapter to the development of a methodology to assess workload? The original objective of this assessment has not changed. We still aim at uncovering the limits of the central processing mechanisms and argue that such a measurement is essential to improved prediction of behavior. What has been changed substantially is the scope and details of the proposed assessment. A global measure of workload does not appear achievable. Instead, any claim or statement about workload should be accompanied by an attempt to identify the sources of load, the mechanisms involved and affected, the levels of performer's experience and the criteria of behavior on the task under which this statement is assumed to hold. Although the generality and caution expressed in the above approach may give the impression that workload assessment is back on the drawing board, this is not the case. In this chapter are enough leads, supported by the results of empirical tests, to direct the measurement procedure to concentrate on the most effective dimensions. Clearly, a detailed analysis of the task in question is the basis of any measurement attempt. We have shown probably the most fruitful questions in this analysis and how each is related to patterns of behavior judged to reflect the "cost" of operation of the central processor. Our

claims. At present, the executive and supervisory aspects of consciousness appear to best correspond in these models to the discussion of attention strategies and allocation policy. The cost of conscious activity seems to be most related to the topics of performance organization and task automaticity. Accordingly, the natural candidates for teaming up with this construct are the attention policy mechanism in Kahneman's (1973) capacity model, the effort and evaluation mechanisms in the cognitive-energetical stage model (Gopher & Sanders, 1984; Pribram & McGuinness, 1975), and the processing activity labelled controlled processes in network models of the system (e.g., Norman & Schallice, 1981; Schneider, 1983).

But, as we know, the conscious apparatus is limited and is easily consumed, as already argued by William James (1890). Indeed, there is an obvious relationship between the processes classified by Schneider and Shiffrin (1977) as "controlled" and consciousness. For a discussion of this analogy see Broadbent (1982), Posner (1978), Logan (1978), Laberge (1981). The major contribution of Schneider's work is not in relabeling the distinction between the conscious (controlled) and the nonconscious (uncontrolled), but rather in demonstrating that it is possible to specify those attributes of a task that would allow transforming its performance from the controlled to the automatic mode.

awareness. The limitations on the scope of consciousness are such that consciously manipulating each of the muscles involved in the act of standing would present us with a "workload" exceeding the capacity of all humans. It makes just as much sense that we can not hope to accomplish within the limited capacity of consciousness the incredibly complex task of searching through memory or determining the grammaticality of our sentences. Indeed, when forced to analyze one sentence explicitly and consciously we may find the task leaving no spare capacity at all. And yet we all generate continually, and relatively correctly, an endless stream of sentences, not evaluated consciously for their grammatical structure. In short, information processing is usually unconscious despite the important role played by consciousness. The implications of this perception to the evaluation of workload measures will become evident as we discuss "subjective" measures of workload.

As a specific structure, consciousness does play a role in most of the models discussed. Broadbent (1958, 1971) identified conscious attention with the limited capacity processor of his model. Welford (1967) has associated it with the operation of his decision mechanisms. The elaboration of these single-channel views, and the introduction of multiple resource frameworks required further qualification of these

well defined and understood in cognitive psychology (see e.g., Carr, 1980; Posner, 1978). While undoubtedly from the point of view of our own experience, consciousness is clearly primary, and we view ourselves as guided by the content of our consciousness, it is remarkable that conscious experience does not play a formal functional role in any of the new approaches to describing information found within the organism. Indeed, when treated at all, consciousness plays the role of a specific functional element (see Donchin, McCarthy, Kutas, & Ritter, 1983). Thus some authors equate consciousness with the content of the short-term, primary memory (e.g., Atkinson & Shiffrin, 1971). Others assign to it the role of an internal programmer, executive, supervisor of behavior, and go on to specify its role in specific control of a variety of processing and execution tasks (e.g., Logan, 1979, 1980; Mandler, 1978, 1983; Posner & Snyder, 1975).

A growing body of experimental evidence demonstrates the obvious role played by all levels of the processing system that never reaches awareness (e.g., Dixon, 1981; Marcel, 1980; Underwood, 1979). The surprise expressed at these findings is quite puzzling, since it is evident that much of the brain's information processing is never available to conscious awareness. The numerous vegetative functions requiring information processing are clearly beyond the ken of

represent the most natural organizing framework within which automatic segments of behavior or action schema develop.

2.8.2 Conscious Control and Allocation Policy

We conclude this section with remarks concerning the role of consciousness in this analysis of the limitations on human performance. For William James, as noted in Section 2.1, the limitations on attention were due to a competition for the possession of consciousness. As analysis of human performance continued this framework unifying attention and consciousness disintegrated. The status of consciousness in current models and its relationship to the study of workload have become quite complex. In effect while consciousness is clearly an important element in all tasks, current models of performance ascribe it only a partial role. It is but one structure deployed in the service of behavior. It would be as much an error to assume that our interest is exclusively in consciousness as it would be to ignore it altogether. Successful performance is as dependent on an operator's conscious application of procedural knowledge as it is on his nonconscious implementation of automated sequences developed over the period in which task expertise was acquired.

The mental activity included under the general categories of conscious experience and voluntary control is still not

Analysis of effort and energetical demand can follow the scheme processed by Pribram and McGuinness (1975) and Sanders (1983). A relevant consideration is the ability of the information intake processes and response activation mechanisms to function adequately with minimal involvement of the central effort mechanism. This question is not unrelated to the analysis of controlled and automatic processes. It is reasonable that a development of direct links between encoding and response activation and reduced dependence of these processes on the coordination of voluntary effort can be achieved with increased automaticity. The energetical pools of arousal and activation that regulate these processes can thus obtain independence through the process of increased automaticity. This line of reasoning is an important bridge between energetical and structural constructs in the study of workload. It can also capture the dynamic properties of the workload phenomena in which a person may act more like a single processor in the beginning of training and gradually develop into a multiple resource mode when processing mechanisms and energetical pools gain sufficient independence. We can further speculate that the structural dimensions and processing stages emerging from experimental research as significant qualifiers of central processor work may also

major components of a task are analyzed and evaluated in terms of the main dimensions that have emerged in experimental research.

2.8.1 Dimensions of a Load Profile

To recapitulate, in any attempt to define and measure workload we must attend to structural and energetic aspects. At the same time we cannot ignore the consistency with which stimuli are mapped to responses and the level of practice in the performance of the task. For example, consider the evaluation of an operator's ability to manually control a device while searching memory for information. Relevant dimensions for the analysis of this combination of tasks may include modes of input and response (auditory or visual presentation, verbal or motor responses, etc.), type of central codes (spatial or verbal), hemispheres involvement, and the nature of the requirement of mental operations (e.g., feature extraction, short-term retention, categorization, etc.) as suggested by Wickens (1983), Gopher, Brickner and Navon (1982), Friedman et al., (1981), Sanders (1983). In addition, one should consider the portion of the task that can rely on existing or developing automatic components (Schneider, 1983) and weigh the degree of dependence on controlled processes.

system. The underlying metaphor is one of a self-organizing communication network which develops to improve the transmission of information within the system.

2.8 The Nature of Capacity Limitations

This review of attempts to model the limitations of the processing system makes it evident that in each class of models the historical pattern is similar. One begins with an attempt to develop a simple model, using a formal metric with an intent to use a unidimensional descriptor of the load imposed on the central processor by a given task. The theoretical and empirical analysis conducted within this framework inexorably reveals that the phenomena are much more complex and multifaceted than initially assumed. The load on the central processor can arise in a variety of ways, all describable to some extent within the framework. But in each case the unidimensional framework gives way to a multidimensional version, one bottleneck becomes several, one serial pathway gives way to cascading stages, and a single unified reservoir of resources is replaced by a multiplicity of such reservoirs. The corollary of this development is that the use of a unique measure of task workload is replaced by a load profile rather than by a single measure or a quantity derived from a unidimensional scale. In a load profile, the

There are two main points at which the theory of automatic processes affects the theory of workload. The first is the claim that different task components may have been automated to different degrees at any given stage of practice, and therefore may impose different demands on the central processor at different stages of practice. The second is the emphasis on a developmental and flexible, rather than an evolutionary and rigid, approach to the emergence of structures in the information processing system.

Evaluation of processing costs based upon an analysis of the degree of task automaticity is clearly orthogonal to an analysis based upon stages of processing, types of codes, or modes of input and output. It emphasizes the consequences of the nature and length of training, and attempts to explain the large variations revealed in the processing demands of the same task between and within individuals based upon their experience (see Anderson, 1981; Laberge, 1975). A systematic consideration of this dimension incorporates the influence of practice in the general model of the flow of information within the organism, without contesting the relevance of other dimensions. The idea of gradual development of automated structures that can be operated as a whole with little investment of processing effort, does present an alternative perspective of the structural organization of the processing

processors, and there is considerable variance among persons in the effectiveness of these processors and in the ability to use them in the service of rather complex tasks. A careful monitoring of workload, as it appears in the operator/task loop, is therefore mandatory to ensure proper system design.

3.2.2 Overview of Section

The remaining pages of this section will consider several classes of measures of workload, assuming that within any task/operator loop the operator can be represented as an ensemble of processing facilities ("structures" or "resources" are other words used to describe the relevant elements of the operator). These are the totality of mechanisms, energetical sources, and limited processors discussed in previous sections. In most tasks the operator must deploy one or more of these facilities and performance is related monotonically to the level of deployment of these facilities. Task difficulty can be expressed in terms of the demands on these facilities by the structural properties of the task. The interaction between task and operator is manifested by the degree to which these facilities can be made available given their inherent limitations. The assessment of these interactions is the assessment of workload.

There are two general classes of measurement techniques: those that focus on the loop and provide some global measure of the interaction and those that attempt to be specific regarding the locus of the interaction. The first body of techniques assumes that despite the structural complexity of the operator and the detailed form in which limitations on the processors can manifest themselves, it is possible to obtain global measures of workload. These approaches are based on theoretical positions discussed in Section 2 under the single-channel, single-bottleneck, or single-resource concept. As seen in O'Donnell and Eggemeier, Chapter 42, such global measures have active and persuasive adherents. In this category fall measures based on subjective judgment and performance-based measures that focus on the performance of the assigned task as an indicator of the workload in the system. There is also a class of "physiological" measures, based primarily on arousal, that have been used as global measures of workload.

The second category includes a variety of procedures designed with an interest in diagnostic (Wickens, 1984) measures of workload. Here the commitment is to view the information processor as a complex structure, with emphasis on its stages, whether serial or cascading. There is a commitment to multiple resources in this class of measures and

to an interest in the fine structure of the interactions. We will review in this category a family of secondary task techniques, beginning with an analysis of the conceptual basis of secondary task measures, and outline some of the advantages and disadvantages of the procedures. We shall then discuss the use within the secondary task paradigm of what are often considered "physiological" measures--the Event-Related Brain Potentials (ERPs). Within the context of workload assessment the ERPs are used as a form of nonovert behavior and they are logically equivalent to other measures of secondary task performance. The key advantage of these covert responses is that they allow an assessment of workload in domains that are opaque to assessment by more traditional techniques.

3.3 Subjective Measures

Subjective measurements of workload are made whenever subjects are asked for a direct estimate of the workload they experience during the performance of a task. The term workload is rarely used when instructing the subjects; instead they are usually asked to report the "difficulty" of the task. However, the target is the <u>experience</u> of difficulty during execution of the task. Thus, the subjects are judging the interactions between themselves and the system.

Subjective methods of measurement have gained popularity during the last decade. Indeed, the "Cooper Harper" scale (see O'Donnell & Eggemeier, Chapter 42) which allows pilots to record their impressions of workload is the technique commonly used in the aircraft industry to measure workload.

There are a number of ways by which the subjective measures are acquired. Usually subjects are presented with elaborate rating scales on which they rank the demands associated with the tasks along a wide variety of dimensions. Thus subjects may be asked to indicate the extent to which they felt time pressure, or to evaluate the physical activity, or the mental effort, or task complexity (e.g., Casali & Wierwille, 1982; Hart, Childress, & Bartolussi, 1981; Wickens & Yeh, 1982). In a sense the subjects are performing a task-analysis when they are making these ratings. The questions asked are putatively about the task and components of the task, rather than about workload. The individual rating scales may be combined into an index of subjective workload. The combination can follow different rules. Commonly the investigators employ a multiple, conjoint, scaling technique for combining the measures (e.g., O'Donnell & Eggemeier, Chapter 42; Reid, Shingledecker & Eggemeier, 1981).

Despite its "subjective" aspects this remains a task analysis procedure because it requires a decomposition and a

separate evaluation of the components of the task. Quite a different approach takes as its object of measurement the perceived workload. The subjects' responses are given as a magnitude estimation. This procedure is adopted from classical psychophysics where the psychological dimension is identified as workload while the physical dimension is anchored in the structure of the task, (e.g., Borg, 1978; Gopher & Braune, 1984).

Subjective measures are easy to obtain and they excel in face validity; there is a compelling sense of relevance in a measure that seems to depend directly on the subject's actual experience of workload. Presumably, subjects are keenly aware of the effort needed to cope with task demands. Indeed, panelists in a symposium held in 1979 in which the measurement of workload was examined concluded that: "If the person tells you that he is loaded and effortful, he is loaded and effortful whatever the behavioral and performance measures may show" (Moray, Johanssen, Pew, Rasmussen, Sanders, & Wickens, 1979, p. 105). These brave words are clear enough, but the theoretical and empirical status of subjective measures of workload remains vague. Particularly obscure is the validity of the measures. Advocates of subjective measures appear to assume that face validity is sufficient. Thus, there are a very small number of studies in which actual rather than face

validity is assessed. The degree to which any subjective measures can be used to account for variance in performance remains unknown. Equally obscure is the relationship between subjective measures and psychological processes, and the behavioral phenomena of which they are supposed to be a manifestation.

To appreciate the problems with the use of subjective measures consider the following hypothetical example. You can ask a cab driver in any major city to assign a number describing the workload associated with driving his cab during the rush hour. You may facilitate the cab driver's task by suggesting that the load level of driving during early morning hours, when the streets are empty, equals 10. The driver is likely to ponder for a while and then produce a number. Assuming that this person also plays chess, he can also be asked to assign a number to the workload experienced during a recent game. Again, the driver will generate a number with little difficulty. But what is the meaning of these two numbers? What do they tell us about the processing mechanisms and the degree to which they were pushed to the limit in the two tasks? Are these numbers, so easily gotten, valid and reliable estimates of workload? What is the theoretical and practical significance of a person's ability to compare and

assign numbers expressing the result of the comparison to such divergent tasks as driving a car and playing chess?

3.3.1 Consistency of Subjective Estimates

Any attempt to examine the theoretical and practical basis of subjective estimates must consider a striking aspect of the relevant data, the remarkable consistency with which the estimates are given. Clearly subjects do not assign workload values to tasks at random. On repeated exposures to the same tasks, even if embedded in a wide variety of other and very different tasks, subjects manifest an impressive level of consistency in the choice of numbers to describe the difficulty of the tasks. For example, Gopher and Braune (1984) gave subjects a battery of 21 tasks, varying in input modality (visual, auditory), type of mission (tracking, memory search, dichotic listening, etc.). Fourteen conditions were single task, and 7 were dual task conditions. The whole battery was performed three times, and subjects were asked to give their estimate of load following performance on each task. In addition they were asked at the midpoint and at the end of the experiment to evaluate the load of each of the conditions in the battery in a single instance. The intercorrelation coefficients for the subjective load profile of the 21 tests, in the five rating instances, were all above

0.90. Reliability coefficients of 0.95 and above were also found in a random split (half) of the subjects. Similar levels of consistency were reported by Hallsten and Borg (1975), who compared subjective ratings of items from a standardized intelligence test. Supportive evidence can also be found in Casali and Weirwille (1984), and in Hauser, Childress, and Hart (1982). The consistency of the relationship between the task and the subjective value assigned to its workload are preserved under constrained and well structured rating methods, but also if magnitude estimation techniques are employed, and the subject is generating his values ad lib. In fact, Gopher and Braune, using ratings that subjects made on 21 different tasks, were able to describe this complex set of judgments by fitting a single power function to the data. The roots of this consistency remain to be determined. More important, how is the "workload" indicated by these measures related to the overt performance?

The correspondence between subjective measures and performance measures is quite low, with operators reporting a relatively high workload even though there is no corresponding deterioration in performance. Similarly, an operator's performance may deteriorate without a corresponding increase in workload. Consider a few illustrations. Hallsten and Borg

(1975) required subjects to rate workload in the solution of a problem extracted from an IQ test. They report that the ratings correlated reasonably well with problem solving (R ranged between 0.70 and 0.80). Another example is a study by Bratfisch, Borg, and Dornic (1972), who report high correlations between subjective magnitude estimations of workload and the objective difficulty of problems in the Raven matrices test. A strong association between subjective ratings and performance variations in a difficult manual control task was reported by McDonell (1968). In contrast, many investigators report a dissociation between subjective estimates and measures of performance. Hauser, Childress, and Hart (1982) assigned subjects a tracking task using four levels of difficulty; they also included four levels of difficulty on a memory search task, as well as a time estimation and an auditory monitoring task. In both studies performance was not correlated with the workload ratings despite an organized and consistent ordering of the experimental conditions according to workload and performance. Some investigators obtain intermediate results. They report a correspondence between the subjective estimates and measures of performance on some of their independent variables, yet for other variables there is partial dissociation between the

workload index and performance (Wickens & Yeh, 1983; Vidulich & Wickens, 1983).

It is noteworthy that Gopher and Braune (1984) report a high correlation (0.93) between subjective measures and a task analysis that leads to an index of task difficulty. This analysis, developed by Wickens (1984) is based upon salient features of the task. The index has four dimensions: (a) familiarity of stimulus (e.g., letters are familiar, random dot patterns are not), (b) concurrency of tasks (single vs dual task conditions), (c) difficulty (e.g., number of elements in a memory set, delay in recall), and (d) resource sharing (e.g., same vs different input modality, same or different coding principle). Thus, the subjective estimate captures an objective and external assessment of a task. It does not, however, serve to predict how well, or how poorly, a subject will do. One may infer from this observation that for all its "subjective" trappings the responses that subjects give when asked for an estimate of the difficulty are based largely on a "cognitive" analysis of their knowledge of the task rather than on an assessment of the way they actually interacted with the task.

3.3.2 Theoretical Considerations in the Interpretation and Use of Subjective Measures of Workload

It is of interest to consider the processes manifested by the subjective measures, and to assess their relation to the processes of interest in workload measurement. By their very nature subjective measures depend on the content of consciousness. The operator, after all, is instructed to report the perception of interaction with the task. This report can only be based on those aspects of the interaction of which the subject is aware. Evidently, these measures will serve their purpose only to the extent that the critical aspects of the interaction between the subject and the task are available to consciousness.

There is considerable controversy regarding what is and what is not available to consciousness. In this discussion it is useful to recall the categorization of the contents of consciousness proposed by G. Mandler (1983). Three main categories of events are identified by Mandler: (a) we are conscious as we acquire new knowledge and behavior; (b) conscious processes are active during exercise of choice and judgment; (c) conscious processes exercise an important function during "troubleshooting." This view is reminiscent of the functional relation that Broadbent (1982) proposed

between consciousness and controlled processes, as the term was defined by Schneider and Schiffrin (1977).

If one accepts these classifications of the nature of consciousness, we have to accept that they constrain the information processing activities, variations in which most are likely to be captured by the subjective measures. It seems clear that the measures are likely to reflect those aspects of tasks that require a guided (voluntary) involvement in coping with novelty, generation of new action plans, selection among alternatives, and commitment of the limited capacity working memory. In short those activities that regulate the allocation of voluntary attention, include problems in performance to which the service of this mechanism is called.

These are indeed important components of task performance but the relative share of this range of activities is in the totality of the human information processing system as it copes with the demands of a specific task. Clearly, there is a substantial ensemble of routine, highly practiced, tasks such as driving, speaking, or eating, that use a very small component of processes involving consciousness. At the same time, virtually every task of intermediate complexity includes elements which require conscious activity. It is reasonable to assume that the degree to which the subjective ratings

predict performance is a function of the proportion of controlled, conscious, activities necessary for task performance. This hypothesis explains the fact that subjective measures sometimes correlate with and sometimes are dissociated from performance. This view is consistent with the high correlation found by Gopher and Braune (1984) between subjective measures and Wickens's index of analysis of task difficulty. The four dimensions of the Wickens index (see ips.) may represent those features of the task that attracted the attention of subjects and influenced their voluntary allocation of resources. However, this component of processing is only a fraction of the total processes that determine actual performance. The relationship might have been different with problem solving tasks, or if variations in the standard of desired performance had made the distinguishing attributes among experimental conditions (e.g., Gopher, Navon, & Brickner, 1982; Gopher & Navon, 1980).

It would appear that subjective measures have limited but important functions. They provide converging information, they may help clarify the dimensions of the tasks, and they would be almost entirely sufficient in tasks that depend primarily on controlled processes. But their limitations must be recognized. An alternate view (e.g., Logan 1978; Norman & Shallice, 1983; Schneider, 1984) is that all the effects of

attention should be looked upon as a directed (or controlled) biasing force of limited power, sometimes also described metaphorically as a gain factor. Only those processes that use these biasing factors "load" the information processing systems; the rest are data driven, automatic, and can flow in parallel at no cost to the central processor. Such automatic processes do not limit the information processing system and therefore are not relevant to the measurement of workload. According to this logic, a dissociation between subjective measures and performance is due largely to failures in performance that do not result from limits on the central processor.

3.3.3 Methodological Considerations

The use of subjective measures presents yet another problem. Current techniques of subjective measurement are retrospective in nature. That is, the operators are asked to evaluate the task some time after performance. Thus, even accepting the theoretical validity of these measures, their utility is constrained by the limited capacity of working memory. It is reasonable to assume that a certain portion of the information available to the operator while performing the task is either not available, or may have been distorted, by the time the information is sought. This problem is common to

AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 4/9

UNCLASSIFIED 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10 NL

all measures that depend on verbal reports as data (for a review and theoretical treatment see Ericcson & Simon, 1980, 1984; Nisbett & Willson, 1979). It may very well be that the subjective measurements of a task are determined by some salient features that are compared to general knowledge bases made of the subject's past experience. Those knowledge bases are relevant, but only in a general sense, to the load of the presently performed task. This may be another source of dissociation between performance and subjective measures. There are no data that pertain directly to this issue. We may try to instruct the subjects to produce estimates during performance, but this procedure may constitute a demanding secondary task that will interfere with performance (Ericcson & Simon 1980, 1984).

3.4 Performance Measures, Primary Task

In our discussion of subjective measures we emphasized the lack of correlation between these measures and the manner in which the operator performs the task. Those comments implied that the operator's ability to perform a task ought to serve as a validating measure for any other measure of workload. This attitude is actually quite common. It appears straightforward that the level of performance an operator achieves on any task is a rather direct measure of the

difficulty, and by implication the workload associated with the task. It is recognized that fluctuations in performance may reflect changes in motivation. However, in this discussion we assume that neither motivation nor the basic ability change. We are concerned only with changes in performance due to limitations on the information processing system. In this case it would appear natural to use the performance on the primary task as a measure of workload.

We will not review here specific measures of primary task performance. Their implementation is fairly obvious and the reader is referred to O'Donnell and Eggemeier, Chapter 42, for more detail. This section focuses on the factors that reduce the utility of direct measures of performance in the study of workload.

When the difficulty of a task is increased, more resources are required by default to maintain the same level of performance. If these resources are available, performance may remain unchanged. Hence, even though no change in behavior is observed, workload has increased and therefore the evaluation of observed "direct" measures is quite different.

Consider a study in which the speed of a target is increased for a subject tracking the movement of a target on a computer screen with a dual-axis hand controller. We can expect an increase in target speed to be accompanied by an

increased demand for processing and response efforts. If tracking accuracy is not changed, we can assume that additional resources were indeed deployed to maintain this level of accuracy, but have no clue from the behavior itself as to how much and how loaded the new tracking condition is. If an increase in tracking error is observed, we still do not know whether those decrements reflect all the "costs" or only part of the "costs" to the system (see Gopher & Navon, 1980; Navon, Gopher, Spitz, & Chillag, in press, for a discussion of attention demands of tracking tasks). Moreover, under the multiple resource notion, decrements may result from exceeding the limits of one or several different processors. We usually have little indication in direct performance measures as to which of the underlying capacities has been overloaded.

A complementary problem to the evaluation of task difficulty manipulation is the scaling approach to map changes in required standards of performance to amounts of invested resources. Logically the marginal costs of every additional unit of performance will not be the same across the whole range of task performance, and a greater effort per unit of performance will be demanded at higher levels. We follow Norman and Bobrow (1975) and Navon and Gopher (1979) in proposing the idea of a hypothetical Performance-Resource-Function (PRF) that relates actual performance to the amount

of invested resources and the suggestion of changing rather than fixed costs for additional units of performance. Within this framework, comparisons among the resource costs of different levels of performance should consider their location on the PRF. In the example of the tracking task, if a subject is able to comply with our increased accuracy requirement, we can assume that it is done with additional costs to the system, but we are still unable to determine the amount and nature of these additional costs. If performance does not improve, it is again unclear whether no additional resources can be recruited to tracking performance, or that performance reached an upper ceiling such that additional resources would not improve accuracy.

In summary, direct measures of performance on the task of interest are usually a poor indicator of mental workload because they often do not reflect variation in resource investment due to difficulty changes; they do not diagnose the source of load, and they do not enable a systematic conversion of performance units into measures of relative demands or load on the processing system. The attempts to develop a workload measurement technique based upon performance measures have been almost exclusively limited to the strategy of studying patterns of interference in the concurrent performance of tasks under dual task conditions. There are two main variants

of this experimental paradigm, the secondary task technique and the Performance Operating Characteristics (POC) methodology (see also O'Donnell & Eggemeier, Chapter 42, and Sperling & Dosher, Chapter 2).

3.5 Arousal Measures

A class of direct measures that we shall mention briefly in this section is the ensemble of psychophysiological measures. These measures are obtained by recording, in general non-invasively, signals generated by the activity of some bodily system. Beatty and his colleagues, for example, have reported that the diameter of the pupil is particularly sensitive to variations in mental effort (Beatty, 1979). Various cardiovascular measures of effort have been examined, and advocated, by numerous investigators (e.g., Mulder, 1979; see also Hart, in press). O'Donnell and Eggemeier, Chapter 42, review this literature in some detail.

In general, the assumption underlying this work is that as the demand for mental effort increases various bodily systems are activated, or "aroused," in the process of marshalling resources in the service of this increased effort. This arousal may be manifested through increased cardiovascular activation. It may also be manifested through activation of the parasympathetic system so that pupil

dilation is evident. The signals are readily recordable and can be obtained with minimal disruption to the performance of the task. The recording of cardiovascular activity requires merely the attachments of a few electrodes to the body. Pupillary activity can be recorded without any attachments to the body, though the equipment required to monitor the pupil may be somewhat cumbersome (see Hamilton, Mulder, Strasser, & Ursin, 1979).

The validation of these "physiological" measures of Workload has been based on the recording of changes in the measure under conditions in which control variation of workload is induced. Beatty's work has been particularly elegant as he has been able to exercise rather tight control over the pupillary diameter by varying the cognitive load imposed on the subject. Thus, the pupil will dilate in a rather precise manner as a function of the requirement to perform mental arithmetic, or as a function of the selective attention demands of the situation.

A key problem in the interpretation of these data, and in the utility of these measures of workload is related to the specificity of the response. In a way, the problem is quite similar to the problem one encounters in the attempt to use psychophysiological measures as indices of deception. It is quite clear that the stress associated with deception does manifest itself in an ensemble of observable physiological

changes reflecting changes in "arousal." It is however equally clear that exactly the same changes may be observed in connection with stress caused by many factors other than deception. Thus, it is not the physiological measure per se, but rather the context within which it is recorded that determines the value of any psychophysiological measure. The investigator must create a setting within which the physiological changes, and the arousal they signify, can be interpreted in an unambiguous fashion. It is for this reason that we are not entirely persuaded of the value of global measures of arousal in the assessment of workload. A discussion of the manner in which psychophysiological measures can be used in a more specific manner is provided in Section 3.9.

3.6 Specific Measures

It is not surprising that approaches to workload measurement based on the assumption that workload is an entity that globally characterizes the interaction between an operator and a task have been found wanting. We have reviewed the many attempts to define the concept and to map the limitations on the human information processing paradigms. The conclusion seems clear. Because the information processing system is diverse in its structure, it can go about its tasks in many different ways. Multiple and not altogether

predictable strategies may be used by different operators under different circumstances, for different tasks. This diversity allows system limitations to appear in different guises and to be circumvented by operators who will employ all handy stratagems to ensure that they cope with their assigned tasks.

Once the system is so conceived, it is natural to examine the possibility that measurement of the limitations would address specifically the full repertoire of possible limitations. If workload is a consequence of a diverse set of interactions taking place in different "structures," it is critical that workload be measured in a structure-sensitive manner. Measures must therefore be designed to allow monitoring of the different structures whose stress under excessive demands brings forth the increase in workload. We review here two closely related approaches, both from the "secondary task" technique.

In this measurement paradigm the workload associated with a given task, the "primary" task, is measured by assigning the operator another task to perform concurrently with the primary task. The operator is told that this new task is "secondary" in importance. The primary task must be performed to the best of the operator's ability, even if this means neglecting performance on the secondary task. Fluctuations in

performance of the secondary task are therefore assumed to reflect fluctuations in workload associated with the primary task. An assumption underlying this technique is that performance on any task depends on the measure of the resources allocated to that task. It is further assumed that the pool of resources is fixed in magnitude. From these two assumptions it follows that there is a reciprocal relationship between the performance in each of the two tasks. Deterioration in the performance of the secondary task must be due to an increase in the resources drawn to maintain the level of performance on the primary task. So presented, this class of measures assumes a general, undifferentiated pool of resources. It is possible, however, to extend this logic to a system characterized by pools of specific resources by selecting secondary tasks assumed to draw from one or another specific pool (Knowles, 1963; Rolfe, 1973).

3.6.1 Selection of a Secondary Task

Thus the selection of a secondary task depends on the view adopted regarding the nature of the central processor. Under the single-capacity, undifferentiated resource models that dominated the theoretical thinking during the 50's and 60's (Broadbent, 1958; Kahneman, 1973; Moray, 1967), it was logical to search for a single standard secondary task that

can constitute a common ruler along which all other tasks can be scaled and compared. The hope to find such a task guided the work of Michon (1966) when he proposed his tapping task (Michon & von Doore, 1967); the research with critical tracking tasks conducted by Jex (1967, 1976; Jex, et al., 1966), and the development of a workload index based upon pupil dilation as attempted by Kahneman (1973). However, it has proven impossible to develop a standard task. It was particularly difficult to find a task whose assignment will not interfere with the performance of all primary tasks. It has also become increasingly clear that tasks are selectively sensitive to interference by one group of tasks and indifferent to members of other groups; this fact was an important contributor to the refutation of a single capacity view (e.g., Gopher, Brickner, & Navon, 1982; McLeod, 1977; Wickens, 1976). These findings account for the predominance of the "structural" approaches to the measurement of workload. Within the structural framework, the selection of a secondary task is much more complex. The experimenter accepts in advance the argument that different secondary tasks can tap different components of the task under consideration. The demands imposed by some secondary tasks can be totally uncorrelated with those of the task in question. The burden is on the experimenter to clarify the exact aspect of the

workload being measured. It is then necessary to identify a secondary task that can best tap this dimension. Several authors provided extensive reviews of the secondary task literature which demonstrate clearly the pronounced consequences of such a selection strategy over a wide variety of primary-secondary task pairings (e.g., O'Donnell & Eggemeier, Chapter 42; Ogden, Levine, & Eisner, 1979; Wickens 1980; Williges & Wierwille, 1979).

One problem with trying to fit the selection of a secondary task to probe the processing limit of interest is that we do not have a good model of the central processor. Hence, all hypotheses on the mapping of task dimensions and performance requirements to the limit of this processor are highly speculative at this time. Favorite dimensions are modality of stimuli, mode of response, coding principles (spatial/verbal), memory load (both working memory and long term memory), time characteristics (self paced/externally paced, continuous/discrete), and level of practice. A second problem arises from the fact that every task, whether primary or secondary, is a compound of demands that can be described along many dimensions. It is impossible to construct a secondary task that taps only a single component as it is unreasonable to assume that a primary task depends only on one kind of processing facility. An attempt to study the demand

structure of a task employing a secondary task procedure is, therefore, likely to require a series of tests each including a slightly different primary-secondary task pairing. These variations are directed towards partialling out structural issues, timing overlap, rate of presentation, performance criteria, etc. They may concentrate on the characteristics of the secondary task, the primary task, or both. Examples of this logic can be found in Gopher, Brickner, and Navon (1982), Riesberg (1983), and Wickens and Sandry (1983). The original logic of the secondary task technique is maintained within each of the comparisons of patterns of interference among a single pair of concurrently performed tasks. At the same time the researcher examines the emerging pattern across all dual task combinations.

3.6.2 Types of Interference and Lack of Interference

Given the above considerations of selective interference, an important decision for an experimenter when selecting a secondary task, is whether the interest is in maximizing the interference between the concurrently performed tasks, or in searching for a paired task that can be performed in parallel uninterrupted. Maximization of interference appears to be more consistent with the original secondary task rationale, in which the second task is added to saturate the capacity of the

C2, C3 in a); all points outside the area are impossible (e.g., C4 in a). Values in brackets indicate the relative priorities assigned to each task at that point. Note that the intersections of the functions with the X and Y axes are given the values 0.0 and 1.0 respectively. They are assumed to correspond to single task performance levels on the two tasks (i.e., when one task is fully attended to, 1.0, and the other is at 0.0 emphasis). We shall later discuss instances in which this assumption does not hold, that is the intersection points of the POC curve are either lower or higher than the actual level of single task performance.

Insert Figure 41.11 About Here

The three POCs in Figure 41.11 indicate that a POC may assume different shapes that reflect differences in the nature of the tradeoffs between performance on the two tasks. Curve 41.11(a) represents the case of a complete and even tradeoff. Every unit of performance on one task can be traded with performance on the other task, and the "costs" of each performance unit are equal across the whole range of performance on the two tasks. Curve 41.11(c) depicts the opposite case, a complete absence of tradeoff. Each task can progress and regress across its total performance range with

analysis of the behavior of the human processing system, was first introduced by Norman and Bobrow (1975), with reference to a metaphor of a computer system. It was elaborated by Navon and Gopher (1979, 1980), who adopted concepts from microeconomic theory, thus drawing an analogy between a person performing two tasks and a manufacturer trying to optimize his investments in the production of two products. Sperling and Dosher, Chapter 2, use the term Attention Operating Characteristics (AOC) to label these curves and discuss them in the general framework of the theory of signal detectability. Irrespective of the background metaphor the technique employed by all researchers is the same: subjects are instructed to perform graded changes in the relative priorities of tasks under dual task conditions (different emphasis levels are indicated verbally, and often augmented by feedback information and differential rewards). Our discussion is based mainly on the terminology and arguments developed by Navon and Gopher (1979, 1980).

Figure 41.11 depicts several hypothetical POCs describing the possible tradeoffs in the performance of task X and task Y. Each POC traces the bounds of joint performance, and can thus be considered to represent the production frontier line to follow economics terminology. Each combination on this curve or inside the area bounded by it is feasible (e.g., C1,

little capability to disambiguate the reasons for these changes. (c) The paradigm does not enable the assessment of problems related to attention control and strategic planning. The next section examines the ways in which these problems are approached in the Performance Operating Characteristics (POC) methodology.

3.7 Performance Operating Characteristics

A POC is a curve depicting all possible dual task combinations arising from splitting a common and limited pool of resources among concurrently performed tasks. Given the structure of tasks and the capabilities of the system, some levels of joint performance are feasible while others are not. A POC depicts the set of all dual task combinations produced when the system operates at its full capacity. The term full capacity is used to refer only to those facilities competed for by a pair of concurrently performed tasks. The overlap in processing demands between tasks may be partial, and some resources may be relevant to the performance of one task only. That is, a POC is a performance tradeoff function which describes the improvement of performance on one task due to added resources released from lowering the standard of performance on another task with which it is time shared. The idea of constructing POCs, as a general approach to the

lead to the expected outcomes. But what if some levels of allocation are mandatory and imposed by the tasks themselves? How much control do subjects have on the allocation of their processing facilities? Can they detect deviations from optimal allocation? It is clear that these questions are an integral part of the measurement paradigm, as much as the selection of secondary tasks and deciding about the types of difficulty manipulations. Moreover, they need to be examined in a broader theoretical perspective, because they influence the degrees of freedom that the processing system has in coping with situational demands (Moray, Chapter 40; Navon & Gopher, 1979, 1980). Several experimental works have shown that attention control is a skill that can be acquired and improved. It was also shown that without training subjects tend to converge on suboptimal solutions (Gopher, 1981; Gopher & Brickner, 1979; Gopher & North, 1977). This issue deserves explicit treatment.

The main problems of the secondary task methodology can be summarized as follows: (a) single to dual task performance comparisons are risky because of emerging properties and structural changes; (b) the primary task protection requirement is highly susceptible to subjective interpretation. It is often violated and experimenters are required to deal with performance changes on both tasks, with

performance be interpreted? The subject may not have clear answers to these questions. The introduction of a secondary task complicates the situation. Since a second task is introduced, it is clear that some performance on it is expected. This expectation however conflicts with the instruction to fully protect primary task performance. Subjects must decide how much to protect, when is the protection requirement satisfied, and how much effort (or "resources") can be devoted to the secondary task. The subjects may be motivated to maximize their secondary task performance because this presents a real challenge. They may be driven to adopt a lenient interpretation of the instruction, or sacrifice aspects of performance that are not directly monitored by the experimenter (for example, reduce their accuracy when only speed is monitored, or limit their efforts to store information in memory for future reference. See also Sperling & Dosher, Chapter 2). It may also be the case that subjects do not have self knowledge, or good external feedback on their performance and thus do not know what to protect.

Another implicit assumption of the instructions is that subjects have sufficient control on their processing efforts, such that they are able to determine for each pair of tasks how much to give, and readjust their policy if it does not

3.6.4 Allocation Policy

If the introduction of a secondary task fulfills its purpose, the capacity of some processor is saturated and a state of shortage in processing facilities is created. The subject is confronted with the need to develop an allocation policy. That is, the subject must decide how to divide the scarce resources. The subject's abilities to control this allocation policy, and the degree to which options are open to the subject, are aspects of the secondary task paradigm that should not be ignored. The common approach has been to designate one task as primary and ask subjects to protect its performance, thereby creating a priority difference between the concurrently performed tasks. In previous paragraphs we described some of the difficulties that may prevent subjects from protecting primary task performance. These difficulties may result in an unavoidable impairment of performance on both the primary and the secondary tasks, leaving the experimenter with the problem of weighting decrements on one task and improvement on the other. But there are further complications in the implementation of this paradigm.

One problem is that the instructions are rather vague. The precise meaning of primary and secondary may be unclear to the subject. How should the efforts be allocated between tasks? How should the instructions to protect primary task

tracking task with a single display and a single hand controller, subjects can easily trade tracking along vertical and horizontal axes (Gopher & Navon, 1980; Navon, Gopher, Spitz, & Chillag, 1984). Is it then a single- or a dual-task situation? There is no good answer to this question, because there does not exist a clear definition of a task. Note however that an acceptance of the spirit of this argument changes the general locus of emphasis of the methodological approach. While the main focus in the original approach was on the difference between single task and dual task performance levels, our interest now concentrates upon the dual task situation and we compare single to dual tasks only as a secondary source of information. This is the approach proposed by Kantowitz and Knight (1976), and is strongly advocated by Navon and Gopher (1979) and Gopher and Sanders (1984). The dual task situation is the focal point for these researchers, and manipulation of task variables under dual task conditions is the main vehicle for uncovering the demand structure of tasks. Experimental examples of this approach can be found in Kantowitz and Knight (1979) and in Gopher, Brickner, and Navon (1982).

across locations and not solely by the local stimulus presented at each attended location. In a second experiment, special difficulties were demonstrated when stimulus response mappings were different for two choice reaction tasks performed in close succession [Psychological Refractory Period (PRP) paradigm]. A third experiment showed that when the two hands performed different rhythmic actions (internally programmed sequences of taps), there was some tendency of each hand to carry out the action assigned to the other (see also Kelso, Southard, & Goodman, 1979). In all of these examples, extra costs and considerable interference were added due to factors that did not exist and were not relevant to the performance of the single tasks themselves.

The theoretical consideration is whether we want to eliminate or separate the factors that emerge from the combination of the tasks. Emergent processes reflect the reality of the system attempting to coordinate and overcome the problems of two demanding missions. As such, they should be of as much interest to the researcher of the system limits as are the demands imposed by the performance of each of the composing tasks. This argument is even more compelling when one considers the vague status of our notion of a task. Every task is composed of many elements, and subjects can frequently trade them in their performance. In a two-dimensional

components of a pair. Subjects often integrate the two tasks to reduce demands. The problems encountered in the performance of this new task cannot be easily reduced to the demand profiles of the original combining tasks (Hirst et al, 1979; Neisser, 1976).

Even in the event that task integration did not occur, new factors that increase or temper interference in concurrent performance of tasks may appear as the sole contribution of the dual-task condition and be absent or irrelevant to the processing demands of the component tasks themselves. Imagine that the primary task of interest is a tracking task and the secondary task requires subjects to classify words presented in an adjacent window. The requirement to move the visual fixation point from the tracking display to the word presentation may impose heavy constraints on the ability to time share the performance of the two tasks, but has nothing to do with the load imposed by each individually and the variables that can be used to manipulate the processing difficulty on each. This is a rather clear example; others may be harder to detect or account for. Duncan (1979) brings several experimental examples of the effects of such emergent properties on dual task performance. In one experiment the perception of letters in each of several attended spatial locations was shown to be affected by properties combined

reaction tasks were studied jointly with two levels of difficulty of a mental arithmetic task. Another example is a recent study by Gopher, Brickner, and Navon (1982), in which two versions of a typing task were combined with a tracking task under three different levels of desired performance.

3.6.3 The Problems of Concurrency

Within the secondary task methodology, the secondary task is of no interest by itself; it serves only as a tool to enable the study of the demand composition and attention requirement of a single task of interest. Its employment includes an additional implicit assumption that the basic structure of these requirements remain unchanged. This implies that the demands on the operator's resources imposed in the dual task conditions are a linear summation of the demands imposed by each individual task when performed by itself.This "task invariance" assumption comprises two components: (a) when combined the two tasks do not change their nature, and (b) the dual task situation does not have different or emergent properties not present in the single-task condition. As has been argued by several authors the validity of these assumptions is rather uncertain (Duncan, 1979; Neisser, 1976). With training under the dual task conditions, many tasks are not viewed as independent

with asynchronous rhythms. These factors may all create large differences in our ability to time share the performance of tasks. The structural label extends to the typology of processing stages (e.g., Gopher & Sanders, 1984; Sanders 1983), or to the sequence at which relevant processing structures are employed (e.g., Kerr, 1983; Triesman, 1969). Energetical concepts have been introduced to describe interference in those cases in which the structure of the tasks remains unchanged but the intensity of involvement of one or all mechanisms is manipulated. This is usually done by increasing the difficulty of one task variable, say the speed of movement of a target in a tracking task (Gopher & Navon, 1980), or using several levels of difficulty in a number counting task (Riesberg, 1983). A second technique is to change the level of required performance on a task (e.g., the speed of response, or the tolerance level for errors (Gopher, Brickner, & Navon, 1982).

Operationally the two manipulations are translated into qualitative and quantitative changes of the properties of concurrently performed tasks. Because the results of such changes are regarded as complementary rather than conflicting, many studies include a manipulation of both. One example of this approach is a study of McLeod (1977), in which the effects on tracking of two structurally different choice

tasks, integrate them if possible, and maximize their overall performance. Experimental examples for this approach are a study by Allport, Antonis, and Reynolds (1972) in which subjects were required to time-share the playing of one musical piece on a piano and the reading of another piece; a study by Hirst, Spelke, Reaves, Charack, and Neisser (1979), in which subjects read text presented on a screen while writing down dictations presented orally; and a study by Schneider and Fisk (1982), in which subjects performed a category classification and a choice reaction task simultaneously. In all of these studies initial interference disappeared after prolonged practice. The experimental approach and the interpretation of data in these studies have a different focus. The aim of the workload analysis should therefore be specified and stated clearly in advance.

Another concern in generating and interpreting patterns of dual task interference is related to our distinction between structural and energetical dimensions of processing and response limitation. When the effects of vocal response are contrasted with manual response and found to be different, they are attributed to a structural factor, namely, the mode of response. Similarly we regard as structural the inability of the eyes to move fast enough to monitor cnanges on a multi-instrument display, or the problems of coordinating actions

assumptions are violated. The hope that the performance on the primary task will be protected and the secondary task will tap a relevant common dimension, has turned out to be rather naive. To uncover the demand composition of a task it seems more realistic to start with a situation where the existence of a strong interference between tasks has been demonstrated for both tasks, and work our way through the origins of this interference.

A different perspective of the dual task paradigm is emerging when the main objective of the measurement is the ability to perform, in parallel, two tasks with minimal interference. Here, the main thrust is not a study of the processing demands of one of the tasks in the pair, but rather on the global efficiency of the dual-task situation. To phrase it differently, how many things can be done simultaneously? In the selection of tasks the effort is to minimize the overlap in demands and eliminate interference. In principle, this approach is complementary to the secondary task technique, but it marks both a strategic and a theoretical shift in emphasis. Theoretically, a lack of interference is less amenable to association with a specific dimension and should be considered as an overall integral property of the situation. Strategically subjects are encouraged to put equal emphasis on the performance of both

system, create an overload, and enable one to scale the demands of the primary task. It is, therefore, somewhat surprising that a lack of obtrusiveness of the introduction of a secondary task to the performance of a primary task has been identified by several authors as a highly desired property of a good secondary task (e.g., Csali & Wierwille, 1982; O'Donnell & Eggemeier, Chapter 42; Williges & Wierwille, 1979). How can this aspiration coexist with the main thrust of a technique that advocates the study of interference patterns as its main tool? There appears to be an additional, implicit assumption by these authors. It is assumed that performance on the primary task can be completely secured when the secondary task is introduced, such that all consequences of the overload would show up only in secondary task performance. This is, of course, a very convenient state of affairs from a methodological viewpoint, but it also requires the assumption that: (a) subjects are in full control of the allocation of their processing efforts among the two tasks (see Navon & Gopher, 1979); (b) the introduction of the second task does not cause an important change in the nature of the performance conditions of the primary task (e.g., Duncan, 1979); and (c) there are no minimal processing costs of the secondary task created by its mere introduction (e.g., Gopher, 1981; Norman & Bobrow, 1975). It is quite likely that these

no effect on the performance of the other task. Both (a) and (c) are demonstrations of a clear relationship easily explained. The more complex and difficult (and more frequent) to interpret is curve 44.11(b), which shows convex reflecting changing rates and uneven processing costs at different regions of the performance range. There are several possible causes for this type of a tradeoff function, and they are not mutually exclusive. One possibility is that one or both tasks have reached a performance ceiling or a data limitation such that resources released from one task cannot be used to improve performance on the other. Another reason is that the marginal costs of performance on each task may be different at different levels of performance and the POC will be sensitive to this change of costs. A third possibility is that tasks overlap only partially in their demand for common resources, and the influence of this overlap on joint performance restricted to one region of performance or change in different regions. The convex curve and its degree of curvature in this case represent an intermediate class between a complete tradeoff and the total independence depicted in Figure 41.11(a) and 41.11(c). An interpretation of a POC may be beneficial, and may observe the considerable broadening of the scope of questions examined within this paradigm relative to the limited range provided by the traditional secondary task

approach. It seems instructive to review a few more features of the POC technique before turning to consider experimental examples.

Recall that a POC is obtained by changing the relative emphasis on concurrently performed tasks holding all other variables at a constant level; it is basically a test of the influence of a change in processing efforts under fixed difficulty conditions. If in addition the variables or the structure of tasks is manipulated, a new POC has to be constructed for every new condition, leading to a family of POCs like the one depicted in Figure 41.12

Insert Figure 41.12 About Here

Quadrant 1 in this figure presents a family of three POCs. They were obtained by manipulating priorities in the joint performance of task X with task Y, holding constant the difficulty of task X and changing the difficulty of Y in three levels (Easy, Medium, and Difficult). It can be seen that as the difficulty of task Y increases, more units of performance on task X have to be sacrificed to improve each unit of performance on task Y. Consequently the curve intersects the Y-axis at a lower point and has a shallower slope. Quadrants 2 and 4 in Figure 41.12 demonstrate how the joint POCs are

related to the individual performance resource functions of each task. The third quadrant depicts the allocation policy line and the influence of two priority levels (A, B) on concurrent performance levels.

The POC methodology has broadened considerably the scope of usage of interference patterns among concurrently performed tasks as a method for studying workload. Its major assertion is that to analyze the demands of a pair of tasks, one cannot use just one condition in which demands are imposed, as this would be considering a single point, and at an unknown location, from a complete curve. An analogous problem would be the selection of a single point on an ROC curve in signal detection theory. The limits of such an approach are quite obvious from examining the hypothetical POCs in Figures 41.11 and 41.12 (for a further discussion of this point see Navon & Gopher, 1979, 1980; Sperling & Dosher, Chapter 2). The adoption of the POC method entails a systematic consideration of allocation policy effects and motivational factors neglected by the traditional secondary task paradigm. The value of collecting these data is not only to control the influence of such variables, it also enables a better assessment of the contribution of processing effort (energy investments) to the performance of tasks and the degree to which tasks interfere with each other, holding all other

variables constant. Given the inherent ambiguity of the specification of task structure, the power and importance of such an additional comparison are much increased. *This is* especially so if a family of POCs is constructed by manipulating variables of task difficulty.

The main drawback of the approach is that it is considerably more complex in the design of experiments. It is also more time consuming in data collection and analysis. However, when one considers the richness and importance of the information, and the present state of our knowledge of workload, the additional effort seems worthwhile. We next review one example from a paper by Gopher, Brickner, and Navon (1982) to demonstrate the application of this approach. The reader can find another example in Sperling and Melchner (1978; see also Sperling & Dosher, Chapter 2).

To test the notion of multiple resources, a two-dimensional pursuit tracking task was paired with a letter typing task. The relative priorities of the two tasks under dual task conditions and the difficulty of the typing task were manipulated. Each task was also performed singly. A schematic diagram of the experimental display is presented in *Figure* 41.13.

Insert Figure 41.13 About Here

In the tracking task, subjects controlled the movement of a control symbol on the screen via a two-dimensional hand controller with a relatively difficult control dynamic. Target movement was driven by a random forcing function governed by the computer. The typing task required the subject to type the appropriate chord combinations of Hebrew letters presented within the moving target of the tracking task. This task was based on a single hand chord typewriter developed by Gopher and Eilam (Gopher, 1984). The system comprises three keys and each letter is entered by typing two successive chords of one to three keys pressed together. Letter codes provide spatial mnemonics corresponding to the shape of letters in print. Difficulty on this task was manipulated in two ways. Under the cognitive manipulation the number of letters in the set presented to the subject was increased from 4 to 16. In the motor manipulation a group of 4 letters coded by motorically difficult chord combinations were selected. The difficulty of the tracking task remained constant under all conditions. Priorities of tasks were varied through a feedback display composed from a static vertical line and two moving bar graphs, each representing one

task (Fig. 41.13). The vertical line represented desired performance. When it was moved away from the side of one task performance, demands on it were increased while simultaneously decreasing on the other task. Priority changes were thus translated into commensurate increases and decreases of performance demands along the POC curve assuming that subjects operate with full capacity. Desired performance levels were computed relative to a standardized level representing top effort obtained in single task conditions. The moving bar graphs were continuously updated, based upon a running average, and reflected the momentary difference between actual and desired performance on each task. Three levels of priorities (.3, .5, .7) were experimented under dual task conditions in addition to single tasks considered a 1.0 level on the priority scale.

The main findings of this experiment are presented in Figures 41.14 and 41.15. Figure 41.14 depicts the effect of the difficulty and priority manipulations on typing performance. Figure 41.15 describes the family of POCs resulting from the joint performance of tracking and typing under all experimental conditions.

Insert Figures 41.14 and 41.15 About Here

From examining Figure 41.14 it is clear that the two manipulations of typing difficulty and the change of priority levels had marked effects on typing performance. Response times to enter letters decreased monotonically with an increase in the relative priority of the typing task. Note that single task levels lie roughly on the same line. There was also a step increment in response time as a result of increasing either motor or cognitive difficulty. However, only motor difficulty interacted with priority changes and had a steeper slope with response time as task priority varied, relative to the effect of priority change on the easy task version. The same pattern of results is present in the POCs family plotted in Figure 41.15. This figure also shows how performance changes on the typing task were accompanied by commensurate changes in tracking accuracy. These results were interpreted to indicate that tracking and typing compete with each other for motor resources but not for cognitive resources.

The existence of interaction between motor difficulty and priority change and the absence of such interaction when cognitive load was manipulated, were crucial for the above interpretation. If one member of a concurrently performed pair of tasks is made more difficult on a dimension that taps a resource for which the two compete, the slope of the POC is

predicted to change, because a larger sacrifice of performance will be required on the other task to enable an improvement on the now more difficult task. If however, the increase of difficulty is made on a dimension that is not relevant to joint performance, performance on the task on which the manipulation was conducted may be affected, but the slope of the POC should not change because the load on the shared resource has not been changed. A differential diagnostic of the type demonstrated in this experiment cannot be achieved with the conventional secondary task approach. They are demonstrative of the power of the POC methodology. A second benefit of this approach is the ability to compare the relative sensitivity of performance to reallocation of efforts with those that result from a manipulation of the characteristics of tasks.

3.8 Basic Assumptions of the POC Methodology

The POC approach does make a number of implicit assumptions as do other techniques. The first and major one is the sensitivity of performance to processing efforts. Performance on tasks is assumed to be monotonically related to the amount of invested resources. It is further assumed that this amount is the main source of variability in time sharing performance. Problems that arise from data limitations or

performance ceiling (e.g., Norman & Bobrow, 1975), are acknowledged but considered to be only a secondary source of variation. A related assumption is that the human is able to control resources and apportion them at least across the main portion of the response sensitive range. A third important assumption is that the POC reflects the boundary of system performance at full capacity, and that this capacity is fixed. That is, there is a fixed upper limit on the rate of recruitment of processing facilities. If subjects do not operate at maximum capacity, or if capacity can expand and shrink (e.g., Kahneman, 1973), the interpretation of a POC is impossible.

Two additional assumptions are independence of tasks and process invariance. A manipulation of relative priorities under dual task conditions is meaningless unless the component tasks maintain their independence, in the sense that resources allocated to the performance of one are withdrawn from the other. Along the same vein, it is assumed that when the levels of emphasis are changed, tasks do not change their basic processing structure. Only the level at which these structures are engaged is assumed to vary. If the demand composition of tasks changes with the level of allocated efforts, what hopes may one have to find general rules to relate the properties of tasks to processing demands? The POC

methodology, like the secondary task approach, is vulnerable to the effects of emergent properties and task integration. However, it is better equipped to detect and isolate the effect of such variables, in the context of a family of POCs. This is another justification yet for the construction of a family of tradeoff functions. As we define workload in terms of the interactions between operators and tasks, and as the POC curves portray these interactions, their value within a framework is evident.

3.9 Psychophysiological Measures of Workload

Note that while the POC curves do provide a very useful analytic tool of the limits on human capacity, the straightforward structure of the tests employing secondary task methodology is abandoned. This is due largely to the fact that the secondary task may interfere with the primary task. There is thus a need for secondary tasks that do not interfere, or that interfere only slightly, with the primary task. A category of such tasks has been proposed, and developed by Donchin and his colleagues (Donchin, 1975; Donchin, Kramer, & Wickens, 1983; Donchin, 1984), by using Event-Related Brain Potentials (ERPs) as the source of the data from which variations in primary task workload can be inferred. The ERP, and in particular the ERP component called

the P300, is recorded off the scalp of an awake subject and its generation requires no overt action by the subject. The next section introduces the ERP and reviews the way it is used in studies of workload.

3.9.1 Introductory Comments on the P300 Component

The ERP is a transient series of voltage oscillations in the brain that can be recorded from the scalp in response to the occurrence of a discrete event (Donchin, 1975). The ERP is viewed as a sequence of components commonly labeled with an "N" or a "P" denoting polarity, and a number which indicates their minimal latency measured from the onset of the eliciting event (e.g., N100 is a negative going component which occurs at least 100 msec after a stimulus). Since ERPs are small, relative to the ongoing EEG, their study became practical only after the development of reliable signal averagers. These capitalize on the fact that the ERP is, by definition, time-locked to the eliciting event.

It is crucial to recognize the componential nature of the ERP. The effects of the experimental manipulations tend to be quite specific to a few components and a combination of the measures of the entire epoch may obscure the relevant variance. There is a degree of controversy as to the proper identification and definition of components, (Donchin, Ritter,

& McCallum, 1978; Picton & Stuss, 1980). In this chapter, however, we shall follow Donchin et al.'s (1978) definition of an ERP component in terms of the responsiveness of the waveforms to specific experimental manipulations. A component is thus mapped into a cognitive space populated by psychological concepts such as decisions, expectations, plans, strategies, associations and memories. The subset of elements in cognitive space associated with a particular component thus contributes to the definition of the ERP component.

The specific attributes of a waveform that are examined in defining a "component" are the amplitude, latency, and scalp distribution. It is the sensitivity of these attributes to experimental manipulations that defines an ERP component. Although no reference has been made to the underlying neural source of components, it is generally assumed that a scalp distribution which is invariant across repeated stimulus presentations implies a specific and fixed set of neural generators (Goff, Allison, & Vaughan, 1978). Thus the scalp distribution which is related to the underlying neural population responsible for the generation of the component is assumed to be a crucial defining characteristic.

The ERP components discussed in this chapter are "endogenous" and are distinct from another class of ERPs called "exogenous." The exogenous components represent an

obligatory response of the brain to the presentation of a stimulus. These components are primarily sensitive to such physical attributes of the stimuli as intensity, modality, and rate. The seven peaks or "bumps" which occur in the first 8-10 msec after the presentation of an auditory or somatosensory stimulus are a prototypical example of the exogenous category (Jewett, Romano, & Williston, 1970).

Endogenous components, typically, are not sensitive to changes in the physical characteristics of the eliciting stimuli. On the other hand, these components are very sensitive to changes in the processing demands of the task imposed on the subject. The endogenous components are nonobligatory responses to stimuli. The strategies and expectancies of the subject as well as other psychological aspects of the task account for the variance in the endogenous components. A typical example, and one to which we shall devote the remainder of this chapter, is the P300 component.

This ERP component is elicited by rare, task relevant stimuli. A task in which it is readily elicited is often called the "oddball" paradigm. In a study by Duncan-Johnson and Donchin (1977), using this paradigm, the subject was instructed to count covertly the total number of higher pitched tones in a Bernoulli series. In different blocks of trials the relative probability of the two tones was

manipulated. It can be seen from Figure 41.16 that the amplitude of the P300 increases monotonically as the probability of the stimulus decreases. This occurs regardless of which of the two stimuli is being counted. When the subjects were solving a word puzzle and were not required to process the tones the P300s were not elicited.

Insert Figure 41.16 About Here

Note that the ERPs in Figure 41.16 that were obtained in this "ignore" condition show no P300 at all levels of probability. Thus, the amplitude of P300 is determined by a combination of the task relevance and the subjective probability of the eliciting event. This basic finding plays a crucial role in the use of P300 in the assessment of workload.

The demonstration that P300 is elicited by unexpected, task relevant stimuli led Donchin, McCarthy, Kutas, and Ritter (1983) to suggest that "the P300 is a manifestation, at the scalp, of neural action that is invoked whenever the need arises to update the 'neuronal model' (Sokolov, 1969) that seems to underlie the ability of the nervous system to control behavior." The neural or mental model is continually assessed for deviations from inputs and revised when the discrepancies

require a response, the data of Isreal et al. (1980) have demonstrated that P300 amplitude is sensitive to the perceptual demands of a primary task.

Kramer, Wickens, and Donchin (1983) performed a componential analysis of the demands of controlling higher order systems, well validated in the literature, to impose a greater load on information processing resources (Baty, 1971; Fuchs, 1962). By "order of control" we refer to the number of time integrations of the output of a controller (i.e., joystick) and the output of the system. In a first order, or velocity driven system, a deflection of the joystick corresponds to a change in the velocity of the controlled element. A second order, or acceleration driven system, produces a change in the acceleration of the controlled element proportional to the movement of the control stick. Assuming that P300 amplitude is sensitive to the perceptual aspects of a task, then a reduction in P300 amplitude by higher order control should localize the influence of the order variable at the earlier processing stages.

Figure 41.20 illustrates the subject's task. The target appeared on the screen and moved in a straight line, but at a

----

Insert Figure 41.20 About Here

----

task. The subjects were instructed to monitor a simulated air traffic control display either for course changes or for intensifications of one of two classes of stimuli (triangles or squares). Primary task difficulty was manipulated by increasing the number of elements traversing the CRT (Sperando, 1978). The numerosity variable did have a systematic effect on reaction time to the tones when subjects were monitoring for course changes. Reaction time increased monotonically from the control condition to the condition in which subjects were required to monitor eight elements simultaneously. However, in the flash detection condition reaction time did not increase significantly as a function of the number of elements displayed.

As can be seen from Figure 41.19 the P300 elicited by the counted tones decreased monotonically with increases in difficulty in the monitoring task when subjects were detecting

----

Insert Figure 41.19 About Here

----

course changes. In the flash detection condition P300s decreased with the introduction of the monitoring task, but increases in the number of display elements failed to further attenuate P300 amplitude. This result is also consistent with the reaction time data. Since the primary task did not

Thus, it would seem that hand movements did not decrease the amplitude of the P300.

Another interpretation of the results can be developed within the framework of the multiple resource theory reviewed in Section 2.6. The notion that P300 is sensitive to a specific aspect of information processing is consistent with the data, reviewed above, regarding the relation between P300 latency and reaction time. P300 latency appears to be sensitive to a subset of the processes that determine reaction time. Furthermore, P300 latency is influenced by manipulations of factors which are assumed to affect relatively early, stimulus evaluation processes while being insensitve to changes in variables which produce their effect on the later response selection and execution processes. If the manipulation of the dimensionality and bandwidth of the tracking task demand resources associated largely with response selection and execution processes then P300 amplitude should not reflect fluctuations in performance. On the other hand, if the perceptual aspects of a task were manipulated, the amplitude of the P300 elicited by a secondary task would be expected to covary with primary task difficulty.

Isreal, Wickens, Chesney, and Donchin (1980) tested the latter hypothesis by combining the oddball task as a secondary task with a visual monitoring task that served as the primary

tolerate without exceeding a preset error criterion. The results are shown in Figure 41.18.

Insert Figure 41.18 About Here

Again, P300 amplitude is diminished by the introduction of the tracking task, but increases in the bandwidth of the forcing function did not produce systematic changes in the amplitude of the P300. These results cannot be explained easily within the framework of an undifferentiated capacity theory if we assume that P300 amplitude indexes the demands placed on the subject by the primary task. Increasing the bandwidth clearly affects the performance of overt secondary tasks (McDonald, 1973; Wierwille, Gutmann, Hicks & Muto, 1977). The fact that P300 did not change, even though a dramatic drop in amplitude was observed with the introduction of the task, required explanation.

One interpretation of the results is that the P300 is not sensitive to the processing demands of the task but instead reflects the motor activity required by tracking. This hypothesis was tested by Isreal et al. (1980) who instructed subjects to manipulate a joystick with one hand concurrently with the oddball task. The amplitude of the P300 component elicited by the tones was not affected by the motor demand.

difficulty were manipulated by requiring the subject to track in either one or two dimensions (horizontal and/or vertical). The compensatory tracking task was defined as the primary task. In addition to the tracking, the subjects were also instructed to count one of two tones presented in a Bernoulli series of high and low pitched tones. Control conditions were also included in which the subjects performed each of the two tasks separately.

The data indicate that the introduction of the tracking task drastically diminishes the amplitude of the P300. However, no further reduction in P300 amplitude could be observed as tracking difficulty increased by requiring tracking in two dimensions. Even though tracking difficulty, assessed by Root Mean Square error (RMS), as well as by reaction time to the tones, definitely increased with the addition of a tracking dimension, P300 amplitude did not change. Isreal, Chesney, Wickens and Donchin (1980) conducted a similar study requiring subjects to perform a compensatory tracking task concurrently with a counting task. In this case, however, the bandwidth of the random forcing function rather than the dimensionality of the tracking task was manipulated. The bandwidth was increased gradually until the cursor's speed reached the highest level the subject could

that lie at the core of the usage that can be made of P300 in the assessment of workload.

It was the basic assumption of this research program that the oddball task can be used as a nonintrusive secondary task since the ERP-eliciting tones occur intermittently, are easily discriminable, and do not require an overt response. Another advantage of this procedure is that it could be applied uniformly across different operational settings. In other words, the oddball task could be inserted into virtually any operational setting without requiring modifications in the system associated with the primary task. Wickens, Isreal and Donchin (1977) reported one of the first studies in the series using a compensatory tracking task as the primary task and the oddball paradigm as the secondary task.

Figure 41.17 illustrates the experimental procedures used in this and several other studies to be discussed. The

----

Insert Figure 41.17 About Here

----

subjects sat in front of a CRT and were instructed to cancel computer generated cursor movements by keeping the cursor superimposed on a target in the center of the display. This was accomplished by movement of a joystick mounted on the right-hand side of the subject's chair. Levels of tracking

processing entity that is invoked, inter alia, whenever task-relevant, surprising stimuli are present. The routine appears to be performing a role in the context-updating activities that occur whenever an event calls for the revision of the neuronal model or schema of the environment. This model makes certain predictions regarding the relationship between the recall of stimuli and the amplitude of the P300 they elicit. Such predictions were confirmed by Karis, Fabiani, and Donchin (1984). Also consistent with this model is the demonstration by Klein, Coles, and Donchin (1984) that people with perfect pitch do not invoke a P300 when they make auditory comparisons.

It is noteworthy that the subroutine manifested by P300 is invoked only if the stimuli are associated with a task that requires that they be processed. Ignored stimuli do not elicit a P300. But what if the stimuli are only partially ignored? What if the subject is instructed to perform the oddball task concurrently with another task? Would the amplitude of the P300 reflect the centrality of the oddball task? Would it, perhaps, change with the amount of resources allocated to the oddball task? Clearly, if so the P300 may serve as a very useful measure of the amount of resources demanded by the two tasks. It is this series of questions

the P300. It should be noted that the phrase "stimulus evaluation" is used here to denote all of the processes that precede response selection and execution. It is not implied, in the theoretical position outlined above, that P300 latency is related solely to the detection and encoding of physical stimuli. It is more than likely that the processing that leads to a P300 extends beyond strict "stimulus" evaluation and encompasses all aspects of the situation that affect, in some way, the system's need to update working memory.

The P300 component of the ERP provides a metric for the decomposition of stages of information processing which compliments the traditional behavioral measures. In terms of applications to system design and workload evaluation ERPs used in conjunction with behavioral and subjective measures permit the assessment of stage specific task interference effects. For example, if two time-shared tasks interfere with each other, it is usually desirable to know the locus of this interaction. Only by discovering the stage at which tasks interact can systems be designed which minimize operator workload.

3.9.3 P300 and Perceptual/Central Processing Resources

The studies reviewed above provided evidence that the P300 component is a manifestation, at the scalp, of a

in an additive factors design (Sternberg, 1969). The subject's task was to decide which of two target stimuli, the words RIGHT or LEFT, was presented in a matrix of characters on a CRT. The characters were either presented within a 4x4 matrix of # signs (no noise condition) or in a 4x4 matrix of letters chosen randomly from the alphabet (noise condition). Stimulus response incompatibility was manipulated by preceding the target matrix either with the cue SAME or with the cue OPPOSITE. SAME signaled a compatible response. The cue OPPOSITE indicated an incompatible response; the right hand would respond to the word LEFT and the left hand to the cue RIGHT. Reaction time increased when the command word was embedded in noise and when the response was incompatible with the stimulus. The effect of the two variables on the response time (RT) was additive implying that these manipulations influenced different stages of processing. P300 latency was increased by the addition of the noise to the target matrix, but was not affected by the incompatibility between the stimulus and the response. These results support the conclusion that P300 latency is affected by a subset of the set of processes which affect reaction time. The P300 is elicited only after the stimulus has been evaluated. Subsequent processing required for the selection and execution of the response does not appear to influence the latency of

which occurred with a relative probability of 20%, and unrelated words which were presented with the complementary probability. The average P300 latency was shortest for the first condition, intermediate for the second and longest for the third condition. The more complex the discrimination, the longer the P300 latency. A detailed analysis of the single trials revealed that the correlation between P300 latency and reaction time was larger for the accuracy condition (.617) than the speed condition (.257). Kutas et al. (1977) concluded that the data supported the hypothesis that P300 latency reflected the termination of a stimulus evaluation process while reaction time indexed the entire sequence of processing from encoding to response selection and execution. Thus, under the accuracy condition when response selection is contingent on stimulus evaluation processes, P300 latency and reaction time are tightly coupled. However, when subjects perform the discrimination under the speed instructions the processes of stimulus evaluation and response selection are more loosely coupled and hence the relationship between P300 latency and reaction time is not as high.

Additional evidence bearing on the issue of the P300's sensitivity to the manipulation of stimulus evaluation processes has been obtained in a study by McCarthy and Donchin (1981) who manipulated orthogonally two independent variables

selection and execution, then experimental variables which have a different effect on processing time in the two stages should influence the relationship between P300 latency and reaction time. For example, when subjects are instructed to respond quickly with a low regard for accuracy, their responses are probably emitted without full evaluation of the stimulus (Wickelgren, 1977). On the other hand, if subjects are instructed to respond accurately they are likely to perform a more thorough analysis of the stimuli prior to responding. This analysis leads to the prediction that the correlation between P300 latency and reaction time would vary with the subject's strategies. Specifically, the correlation would be high and positive when the subjects are instructed to be accurate. Low correlations would be observed under speed instructions.

Kutas, McCarthy, and Donchin (1977) tested this hypothesis by requiring subjects to distinguish between two stimuli under both speed and accuracy instructions. In one experimental condition subjects were required to discriminate between two names, Nancy and David, presented on a CRT (with relative frequencies of 20 and 80%, respectively). In a second condition, female names comprised 20% of the items and males names 80%. In the third condition, subjects were required to discriminate between synonyms of the word "Prod,"

Evidence that P300 is determined by the amount of time required to recognize and evaluate a stimulus has been reported by several investigators who employed Sternberg's (1966) additive factors methodology (Ford, Roth, Mohs, Hopkins & Kopell, 1979; Ford, Mohs, Pfefferbaum, & Kopell, 1980; Gomer, Spicuzza & O'Donnell, 1976).

Other investigators, employing different paradigms also report that P300 latency and reaction time are positively correlated when stimulus evaluation time is manipulated. N. Squires, Donchin, Squires, and Grossberg (1977) found that P300 latency and reaction time covaried with the difficulty of auditory and visual discriminations. Furthermore, P300 latency varied with the manipulation of stimulus discriminability while reaction time was influenced by both stimulus evaluation and response selection factors. Heffley, Wickens and Donchin (1978) performed an experiment in which subjects were required to monitor a dynamic visual display for intensifications of one of two classes of targets. P300 latency was found to increase monotonically with the number of elements on the display. Since subjects were not required to make an overt response the differences in P300 latency were attributed to stimulus evaluation processes.

If P300 latency is determined by stimulus evaluation time and is largely independent of the time required for response

that differ in pitch, the stimuli elicit relatively short latency P300s. More difficult discriminations result in increases in the latency of P300.

Assuming that manual or vocal reaction time terminates processing, and that P300 is a manifestation of a process that precedes the response then it would be expected that P300 latency and reaction time should positively covary. This prediction has been supported by numerous studies (Wilkinson & Morlock, 1967; Bostock & Jarvis, 1970, Rohrbaugh, Donchin, & Eriksen, 1974). Other investigations, however, failed to detect a relationship between P300 latency and reaction time (Karlin, Martz, & Mordkoff, 1970; Karlin & Martz, 1973).

Donchin et al. (1978) proposed an interpretation of the processes underlying the P300 which may reconcile these contradictory findings. They suggested that P300 latency is determined by the time required to evaluate the stimulus, but is largely independent of response selection and execution time. The correlation between reaction time and P300 latency would, accordingly, vary as a function of the percent of reaction time variance that is accounted for by stimulus evaluation processes. This percentage would be affected by the strategies employed by the subject. The strategies, therefore, should influence the relationship between P300 latency and reaction time (see also Ritter et al., 1972).

exceed some criterion value. The frequency with which the mental model is revised is based on the surprise value and task relevance of the stimuli. Donchin (1981) also argued that the concept of a subroutine is an appropriate metaphor for the activity of ERP components (Donchin, Kubovy, Kutas, Johnson & Herning, 1973; Donchin, 1975). Thus, ERP components may be associated with specific information processing functions which are activated in a variety of different tasks. In the case of the P300, the "subroutine" may be invoked whenever there is a need to evaluate surprising, task relevant events. This interpretation of the changes in P300 amplitude is strengthened by the evidence that has accumulated in the past decade regarding the factors that control the latency of the P300. As the use we make of P300 in the analysis of workload depends strongly on the theoretical interpretation of the component it will be useful to provide a brief review and interpretation of the latency data.

3.9.2 The Latency of the P300 Component

The peak latency of the P300 component appears to depend on the time required to recognize and evaluate a task-relevant event. The latency ranges between 300 to 750 msec following the presentation of a discrete stimulus. For example, fairly simple tasks calling for a discrimination between two tones

randomly selected angle, in the direction of its exit. The subject had to move the cursor into the neighborhood of the target. The time between the appearance of the target and its acquisition by the cursor is called the "acquisition phase." Acquisition was accomplished by manipulating the two-axis joystick mounted on the right side of the subject's chair. Successful acquisition initiated the alignment phase. The target began to rotate at a constant velocity in either a clockwise or counterclockwise direction. The subjects had to rotate the cursor at the same velocity as the target while also keeping the two elements superimposed. The rotation was accomplished by manipulating the single axis joystick mounted on the left side of the subject's chair. A deflection of the stick to the right produced a clockwise rotation of the cursor at an angular velocity proportional to the angle of deflection; a deflection to the left produced a counterclockwise rotation. Deviation from the initial acquisition criterion for more than 1000 msec necessitated a realignment of the elements. Once the subject decided that all of the criteria had been satisfied and the target and cursor were aligned, a capture button could be pressed and the trial terminated.

It was assumed that the alignment phase would be more difficult than the acquisition phase due to increased

perceptual demands imposed by the requirement to control the additional rotational axis. It was predicted, therefore, that the P300 amplitude elicited by the tones, associated with an oddball task run concurrently with the tracking task, would be larger during the acquisition than during the alignment phase.

The ERP results presented in Figure 41.21 confirm these predictions. The P300 amplitude is attenuated both as a

----

Insert Figure 41.21 About Here

----

function of phase, larger amplitude P300s being elicited in the acquisition phase, and system order, larger P300s elicited during the easier, first order tracking. Other investigators employing a compensatory tracking task have also found a systematic relationship between P300 amplitude and system order (Wickens, Gill, Kramer, Ross & Donchin, 1981). These studies, along with additive factors investigators of manual control parameters, have provided converging evidence that system order has a salient perceptual/central processing component (Wickens & Derrick, 1981; Wickens, Derrick, Micallizi & Berringer, 1980) The results might also be useful in the design and evaluation of complex tracking tasks. If operators are required to perform a tracking task with higher order system dynamics, then concurrently performed tasks

should be designed to minimize perceptual/central processing load. We see here, again, how the ERPs provide data that increase the theoretical depth with which one can draw conclusions about the human information processing system.

3.9.4 P300 and Resource Reciprocity

The studies cited above have demonstrated a robust relationship between P300 amplitude and the allocation of processing resources in a secondary task. P300s elicited by secondary task probes decrease in amplitude with increases in the perceptual/central processing difficulty of primary tasks. As outlined previously, one of the basic assumptions of the secondary task technique is that increases in primary task difficulty divert processing resources from the secondary task. The decrement in secondary task performance is believed to reflect this shift of resources from the secondary to the primary task. Thus, it is assumed that there is a reciprocal relationship between the resources allocated to the primary and secondary tasks. If this assumption is correct, then it should be possible to demonstrate that P300s elicited by task relevant, discrete events embedded within the primary task are directly related to primary task difficulty.

Kramer, Wickens, Vanasse, Heffley and Donchin (1981) conducted an experiment in which ERPs were elicited by task

relevant events embedded within a tracking task. The subjects were required to perform a single axis pursuit step tracking task with either first order (velocity) or second order (acceleration) control dynamics. In this task, the horizontal position of a target was determined by a random series of step displacements occurring at 3 sec intervals. The subject's task was to keep the cursor superimposed on the target. Difficulty was varied by manipulating two variables: the degree of predictability of the series of steps and the system order. In the high predictability condition the step changes alternated in a regular right-left pattern. In the low predictability condition the sequence of step changes was random. The magnitude of the changes was unpredictable in both conditions. The two dimensions of difficulty, system order and input predictability, were crossed to create three conditions of increasing difficulty: first order control of predictable input, first order control of unpredictable input and second order control of unpredictable input.

Three different types of probes were employed as ERP eliciting events. In one condition, subjects performed the tracking task while also counting the number of occurrences of a low pitched tone from a Bernoulli series of high- and low-pitched tones. In the second condition, subjects counted the dimmer of two flashes in a Bernoulli sequence. The flash

appeared as a horizontal bar along the path traversed by the target. In the primary task probe condition, subjects counted the total number of step changes to the left. Two control conditions were also included: one in which the subjects counted the probes but did not track, and a second in which subjects performed the tracking task without counting the probes.

The important findings to note in the data presented in Figure 41.22 are the monotonic relations between the tracking difficulty manipulations and the subject's perceived ratings

----

Insert Figure 41.22 About Here

----

of difficulty, as well as those between tracking difficulty and RMS error. Both the subjective and behavioral indices converge on the same ordering of task difficulty. However, these measures do not provide information concerning the underlying resource structure of the task.

The effect of tracking difficulty on P300 amplitude in the auditory condition provide results consistent with previous research (Isreal, et al., 1980; Wickens, et al.,

----

Insert Figure 41.23 About Here

1980). Thus, in the auditory condition, an increase in the difficulty of the primary task resulted in a decrease in the amplitude of the P300 elicited by the secondary task probes. In the visual condition the introduction of the tracking task resulted in a reduction in the amplitude of the P300. However, increases in tracking difficulty failed to produce any further attenuation. In the step conditions, the amplitude of the P300 elicited by the discrete changes in the spatial position of the controlled element increased with increments in the difficulty of the primary task. Thus, the hypothesis of resource reciprocity between the primary and secondary tasks was confirmed. One final aspect of the step tracking study has considerable potential practical utility. The sensitivity of the P300 elicited by visual steps to resource allocation was observed independent of whether or not the subjects were required to count the stimuli. These data suggest that inferences from the P300 about resource allocation and therefore workload can be made in the total absence of a secondary task requirement, a considerable advantage if workload is to be assessed unobtrusively in real-time environments.

3.9.5 Summary and Conclusions

The investigations reported above demonstrate that the P300 as a secondary task can diagnostically reflect primary task workload variations of a perceptual/cognitive nature, uncontaminated by response factors. The absence of overt response requirements provide it with a considerable advantage over the secondary task, in that the oddball count task is considerably less intrusive.

As a secondary task however, the probe task is not entirely unobtrusive and interpretation of the measures still requires the investigator to make certain assumptions about the nature of the primary-secondary task interaction to make inferences concerning operator workload. For this reason our most recent observations that P300 elicited by primary task stimuli also reflect resource allocation are particularly encouraging to the utility of the ERP as a measure of workload in extra-laboratory environments.

4. EPILOGUE

Our examination of the workload concept began with the frequent statement that workload is a multidimensional, multifaceted, construct. It was unlikely, we suggested, that the manifestations of workload will be captured by one unique, representative measure. Our review of the literature confirms

this conclusion. We have seen how, time and again, the diverse searches for a uniform metric were forced, after initial successes, to admit the overriding complexity and multidimensionality of the information processing system. It is impossible to escape the fact that any operator brings to virtually every task a large repertoire of structures and processes that can be deployed in a very flexible manner in the service of the task. Any measurement technique must be sensitive to these factors. Much of our discussion was concerned with the attempt to clarify the nature of the dimensions along which workload varies. We have tried to explicate the attributes that should be considered in the selection of a measurement procedure.

The general goal of the workload analysis has been defined as the assessment of the processing and response limitations of the human information processing system. Our primary thesis is that these limitations are revealed only through the interactions between an operator and the assigned tasks. We considered the nature of the limitations on two levels. On the more theoretical level we examined efforts to model the limitations on the system by characterizing the invariant, open loop, properties of the human processing system. On a second, and perhaps more practical, level we developed the argument that workload at any specific instant

of measurement should be regarded as the joint, closed loop property of the human and the assigned task.

The discussion of general sources of limitations emphasized the close affinity between the study of workload and the literature that focuses on attention. Investigators of both workload and attention are interested in the limitations that are placed on the central processor, and in the cost of mental operations. In the course of studying these two related phenomena, investigators were forced to develop a detailed account of the energetical and structural characteristics of the central processor. The linkage between the areas of workload and of attention has been neglected in the past. We endeavored to show in this chapter that it will be useful if investigators of both workload and attention become aware of each other's concepts and findings. The goals of workload assessment can be better identified and measured with reference to the structure of the processing system and to the interrelation between its variables, as they emerge within the attention and information processing research. Attention research, in turn, can benefit and be enriched by a deeper consideration of the costs of mental operations and their impact on global measures of performance.

In other words, there is a need for developing a more comprehensive view of the ways in which the system utilizes

its degrees of freedom. This question is recognized as crucial for workload evaluation, but is almost completely neglected in research on attention. The vast majority of experiments on attention concentrate on isolated and specific comparisons within a limited segment of the system. In many experiments, the question of costs is not raised at all, as there is no attempt to interpret the effects of the experimental manipulations within the framework of the subject's effort to cope with task demands.

Matters are equally in need of clarification when we consider the situation in which the measurement of workload takes place. Usually, an individual, or a group, is assigned a task. It has generally been thought that estimate of workload can be derived from knowledge of properties of the task as well as from the capabilities of the performer. This argument has been a second pivot in our discussion. We have emphasized in this chapter the need for a closed loop assessment of workload to reflect three basic features of the measurement situation: (a) the fuzziness and arbitrariness of the task concept, (b) the degrees of freedom available to a person in adopting a specific strategy to cope with task demands, and considerations of allocation policy, and (c) the influence of practice levels at the time of measurement.

The concept of a task in the human performance literature designates a composite entity that can be specified on many dimensions. These include its formal properties (e.g., modality and rate of information presentation, and mode of response), the selective emphasis on different aspects of performance (e.g., reaction time, accuracy, retention, and comprehension), and expectations regarding the level of performance on each of the selected measures. Deficits in performance and levels of workload are described in terms of these specifications. They are used to define the criteria for sufficiency and adequacy in the assessment of performance. The analysis is, naturally, not sensitive to effects that are outside of the defined features, and the estimate of load may change drastically when any of the specifications is changed.

In a similar manner, it is reasonable to assume that the performer may have more than a single way to cope with the complexity of tasks, and that different strategies may yield similar achievements when evaluated by global measures of performance. We described throughout this chapter the richness and flexibility of the central processing activity, and cited data from many experiments to show that the path from stimulus and response can be traversed in a variety of ways. The strategy used by the subject at the time of

evaluation is therefore another determinant to be considered in the analysis of workload.

Finally, the influence of practice should not be ignored. In several sections, we discussed the development of automatic segments of tasks and the reduction of resource costs in a system driven by automatic as compared with controlled processes. The actual level of practice that an operator has at the time of assessment, or the desired level of practice that was specified by the assessor as a part of the prerequisites of the task, are therefore an important consideration in the estimate of workload. This factor may reduce workload dramatically with the passage of time, or increase it instantly when novelty is encountered.

In the discussion of measurement approaches, we emphasized the importance of theory based measures. The underlying assumptions, limitations and major strength of the main classes of measurement procedures were reviewed and evaluated. The limits of subjective measures were addressed in detail, and a cautionary note was sounded against awarding the subjective measures a privileged status due to their compelling face validity and technical convenience. Our recommendation for a measurement procedure follows the inclusive definition of workload concept that encompasses both conscious and nonconscious processing activity. Accordingly,

we favor the use of detailed task analysis procedures to uncover the major components of tasks, followed by a battery of performance based measures designed to evaluate the load on each component.

REFERENCE NOTES

Schneider, W. A distributed processing architecture for attention and skill development. Invited paper, Midwestern Psychological Association Meeting, May 1983.

Laberge, D. Automatic information processing: A review. In S. Long & A. Baddeley (Eds.), Attention and performance IX. Hillsdale, NJ: Erlbaum Associates, 1981. (a)

Laberge, D. Unitization and automaticity in reading. In M. Larsman & H. Hunt (Eds.), Proceedings of the Lake Wilderness Attention Conference, July, 1981. (b)

Laberge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Levison, W. H. A model for mental workload in task requiring continuous information processing. In N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum Press, 1979.

Lindsley, D. B. Emotion. In S. S. Steven, (Ed.), Handbook of experimental psychology. New York: John Wiley & Sons, 1951.

Logan, G. D. Attention in character classification: Evidence for the automaticity of component stages. Journal of Experimental Psychology: General, 1978, 107, 32-63.

Logan, G. D. On the use of a concurrent memory load to measure attention and automaticity. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 189-202.

Logan, G. D. Attention and automaticity in stroop and priming task: Theory and data. Cognitive Psychology, 1980, 12, 523-553.

MacCorquadale, K., & Meehl, P. E. On a distinction between hypothetical constructs and intervening variables. Psychological Review, 1948, 55, 95-107.

Mandler, G. Mind and emotion. New York: John Wiley & Sons, 1978.

Mandler, G. Consciousness: Its function and construction. University of California, San Diego, 1983.

Kerr, B. Preplanning for aimed movements: Disruption from preliminary task. Journal of Experimental Psychology: Human Perception and Performance, 1983, 9, 596-606.

Kinsbourne, M. The mechanism of hemispheric control of the lateral gradient of attention. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and performance V. New York: Academic Press, 1975.

Kinsbourne, M., & Hicks, R. Functional cerebral space. In J. Requin (Ed.), Attention and performance, VIII. Hillsdale, NJ: Erlbaum Associates, 1978.

Klein, M., Coles, M. G. H., & Donchin, E. People with absolute pitch process tones without producing a P300. Science, 1984, 223, 1306-1309.

Knowles, W. B. Operator loading tasks. Human Factors, 1963, 5, 155-161.

Kramer, A. F., Wickens, C. D., & Donchin, E. An analysis of the processing requirements of a complex perceptual-motor task. Human Factors, 1983, 25(6), 597-621.

Kramer, A., Wickens, C. D., Vanasse, L., Heffley, E. F., & Donchin, E. Primary and secondary task analysis of step tracking: An event-related potentials approach. In R. C. Sugarman (Ed.), Proceedings of the 25th Annual Meeting of the Human Factors Society, Rochester, New York, 1981.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197, 792-795.

Laberge, D. Acquisition of automatic processing in perceptual and association learning. In P. M. A. Rabbit & S. Dornic (Eds.), Attention and performance V. New York: Academic Press, 1975.

Kantowitz, H. B., & Knight, J. L. On experimental limited processes. Psychological Review, 1976, 83, 502-507.

Kantowitz, H. B., & Knight, J. L. Testing tapping time-sharing. Journal of Experimental Psychology, 1979, 103, 331-336.

Karis, D., Fabiani, M., & Donchin, E. "P300" and memory: Individual differences in the von Restorff effect. Cognitive Psychology, 1984, 16, 177-216.

Karlin, L., & Kastenbaum, R. Effects of number of alternations on the psychological refectory period. Quarterly Journal of Experimental Psychology, 1968, 20, 167-178.

Karlin, L., & Martz, M. J. Response probability and sensory evoked potentials. In S. Kornblum (Ed.), Attention and performance IV. New York: Academic Press, 1973.

Karlin, L., Martz, M. J., & Mordkoff, A. M. Motor performance and sensory evoked potentials. Electroencephalography and Clinical Neurophysiology, 1970, 28, 307-313.

Kaufman, L. Sight and mind, New York: Oxford Press, 1979.

Keele, S. W. Attention and human performance. Pacific Palisades, CA: Goodyear Publishing Co., Inc., 1973.

Keele, S. W. Behavioral analysis of movement. In V. B. Brooks (Ed.), Handbook of physiology: Section 1: The nervous system, Vol. 2, Motion control, Part 2. Baltimore, MD: American Physiological Society (distributed by Williams and Wilkens), 1981.

Kelso, S. A., Southard, D. L., & Goodman, D. On the coordination of two-handed movements. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 229-259.

REFERENCES

Jex, H. R. Two applications of the critical instability task to secondary task workload research. IEEE Transaction of Human Factors in Electronics, 1967, HFE-8, 279-282.

Jex, H. R. A proposed set of standardized subcritical tasks for tracking workload calibration. In N. Moray (Ed.), Mental workload. New York: Plenum Press, 1976.

Jex, H. R., McDonnel, J. P., & Phatac, A. V. A critical tracking task for manual control research. IEEE Transaction of Human Factors in Electronics, 1966, HEF-7, 138-144.

Kahneman, D. Remarks on attention control. Acta Psychologica: Attention and Performance, III, 1970, 33, 118-131.

Kahneman, D. Attention and effort. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Kahneman, D., Beatty, J., & Pollack, I. Perceptual deficit during a mental task. Science, 1967, 157, 218-219.

Kahneman, D., Peavler, W. S., & Onuska, L. Effects of verbalization on incentive on the pupil response to mental activity. Canadian Journal of Psychology, 1968, 22, 186-196.

Kahneman, D., & Peavler, W. S. Insensitive effects and pupillarity changes in association learning. Journal of Experimental Psychology, 1969, 79, 312-318.

Kahneman, D., & Tresiman, A. Changing views of attention and automaticity. In R. Parasuraman, J. Beatty, & E. Davis (Eds.), Varieties of attention. New York: John Wiley & Sons, 1983.

Kahneman, D., Tversky, B., Shapiro, D., & Crider, A. Pupillary, heart rate and skin resistence changes during a mental task. Journal of Experimental Psychology, 1967, 27, 187-196.

REFERENCES

Hillyard, S. A. Event-related potentials and selective attention. In E. Donchin (Ed.), Cognitive psychophysiology: Event-related potentials and the study of cognition, The Carmel Conferences, (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum, 1984.

Hirst, W., Spelke, E. S., Reaves, C. C., Charack, G., & Neisser, U. Dividing attention without alternation of automaticity. Journal of Experimental Psychology: General, 1980, 109, 98-117.

Hyman, R. Stimulus information as a determinant of reaction time. Journal of Experimental Psychology, 1953, 45, 188-196.

Israel, J. Structural interference in dual task performance: Behavioral and electrophysiological data. Unpublished doctoral dissertation, University of Illinois, 1980.

Israel, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 1980, 17, 259-273.

Israel, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. The event-related brain potential as an index of display monitoring workload. Human Factors, 1980, 22, 212-224.

Jagacinski, R. J., Reppenger, D. W., Ward, S. L., & Moran, M. S. A test of Fitt's law with moving targets. Human Factors, 1980, 22, 225-233.

James, W. The principles of psychology. New York: Holt, 1890.

Jewett, D., Romano, H. W., & Williston, J. S. Human auditory evoked responses: Possible brain stem components detected on the scalp. Science, 1970, 167, 1517-1518.

Gopher, D. & North, R. A. Manipulating the conditions of training in time-sharing performance. Human Factors, 1977, 19, 583-593.

Gopher, D., & Sanders, A. F. S-OH-R OH stages, OH resources. In W. Printz & A. F. Sanders (Eds.), Cognition and Motor Processes. Berlin: Springer-Verlag, 1984.

Hallsten, L., & Borg, G. Six rating scales for perceived difficulty (Report # 58). Institute of Applied Psychology: The University of Stockholm, 1975.

Hamilton, P., Mulder, G., Strasser, H., & Ursin, H. Final report of physiological psychology group. In N. Moray (Ed.), Mental workload: its theory and measurement. New York: Plenum Press, 1979.

Hart, S. G. Theory and measurement of human workload. In J. Zeidner (Ed.), Human productivity enhancement: Cognitive processes in system design. New York: Praeger Scientific. In press.

Hart, S. G., Childress, M. E., & Bartolussi, M. Defining the subjective experience of workload. Proceedings of the 25th Annual Meeting of the Human Factors Society, 1981, 527-531.

Hauser, J. R., Childress, M. E., & Hart, S. G. Rating consistency and component salience in subjective workload estimation. Proceedings of the 18th Annual Conference on Manual Control, Dayton, OH, 1982.

Heffley, E., Wickens, C. D., & Donchin, E. Intramodality selective attention and P300--reexamination in a visual monitoring task. Psychophysiology, 1978, 15, 269-290.

Hick, W. E. On the rate of gain of information. Quarterly Journal of Experimental Psychology, 1952, 4, 11-26.

Goff, W. R., Allison, T., & Vaughan, H. G. The functional neuroanatomy of the event-related potentials. In E. Calloway, P. Tueting, & S. Koslow (Eds.), Brain event-related potentials in man. New York: Academic Press, 1978.

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. Evoked potential correlates of visual item recognition during memory scanning tasks. Physiological Psychology, 1976, 4, 61-65.

Gopher, D. The contribution of vision based imagery to the acquisition and operation of transcription skills. In W. Printz & A. F. Sanders (Eds.), Cognition and motor performance. Berlin-Heidelberg: Springer-Verlag, 1984.

Gopher, D. Performance tradeoff under time-sharing conditions: the ability of human operators to release resources by lowering their standards of performance. Proceedings of the International Conference on Cybernetics and Society, 1981, 609-614.

Gopher, D., & Brickner, M. On the training of time-sharing skills: An attention viewpoint. In G. Goonick, M. Hazeltine, & R. Durst (Eds.), Proceedings of the Annual Meeting of Human Factors Society. Santa Monica, CA, 1980.

Gopher, D., Brickner, M., & Navon, D. Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 146-157.

Gopher, D. & Browne, R. On the psychophysics of workload: Why bother with subjective measures? Human Factors. In press.

Gopher, D., & Navon, D. How is performance limited--testing the notion of central capacity. Acta Psychologica, 1980, 46, 161-180.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopking, W. F., & Kopell, B. S. Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 1979, 47, 450-454.

Fraise, P. La periode refractoire psychologique. Annee Psychologique, 1957, 57, 315-328.

Frazier, M. L., & Crombie, R. B. Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics. (AFFTC-TR-82-5), Air Force Flight Test Center, Edwards Air Force Base, CA, 1982.

Friedman, A., & Polson, M. C. The hemispheres as independent resource systems: A limited capacity processing and cerebral specialization. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 1031-1058.

Friedman, A., Polson, M. C., Defoe, C., & Gaskill, S. Diverting attention within and between hemispheres: Testing a multiple resources approach to limited capacity information processing. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 625-650.

Frowein, H. W. Selective effects of barbituate and amphetamine on information processing and response execution. Acta Psychologica, 1981, 47, 105-115.

Fuchs, A. The progression-regression hypothesis in perceptual-motor skill learning. Journal of Experimental Psychology, 1962, 63, 177-182.

Garner, W. R. Uncertainty and structure as a psychological concept. New York: John Wiley & Sons, 1962.

Garner, W. R. The processing of information and stimulus. Patoman, MD: Erlbaum, 1974.

Duncan, J. Divided attention: The whole is more than the sum of its parts. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 216-228.

Duncan-Johnson, C. C. & Donchin, E. On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 1977, 14, 456-467.

Easterbrook, J. A. The effects of emotion on cue utilization and the organization of behavior. Psychological Review, 1959, 66, 183-201.

Ericcson, K. A., & Simon, H. A. Verbal reports as data. Psychological Review, 1980, 87, 215-251.

Ericcson, K. A. & Simon, H. A. Protocal analysis: Verbal report as data. Cambridge, MA: MIT Press, 1984.

Fairbank, G., Guttman, N., & Miron, M. S. Effects of time comparison upon the comprehension of connected speech. Journal of Speech and Hearing Disorders, 1952, 22, 10-19.

Fisk, A. D., & Schneider, W. Category and word search: Generalization search principles to complex processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1983, 9, 177-195.

Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 1954, 42, 381-391.

Fitts, P. M., & Posner, M. I. Human performance. Belmont, CA: Brooks Cole Publishing Co., 1967.

Ford, J. M., Mohs, R. C., Pfefferbaum, A., & Kopell, B. S. On the utility of P300 latency and reaction time for studying cognitive processes. In H. H. Kornhuber & L. Deecke (Eds.), Motivation, motor and sensory processes of the brain: Electrical potentials, behavioral and clinical use. Netherlands: North Holland Biomedical Press, 1980.

Donchin, E. The use of ERPs to monitor nonconscious mentation. In Hart, S. G. & Hartswell, E. J. (Eds.), Proceedings of the Twentieth Annual Conference on Manual Control, June 12-14. NASA-Ames Research Center,California, NASA CP-2341, 1984, Vol. II, pp. 1-20. (b)

Donchin, E., Kramer, A., & Wickens, C. Probing the cognitive infrastructure with event-related brain potentials. In M. C. Frazier & R. B. Crombie (Eds.), Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics (AFFTC-JR-82-5), Edwards Air Force Base, 1982, 371-387.

Donchin, E., Kubovy, M., Kutas, M., Johnson, R., & Herning, R. I. Graded changes in evoked response (P300) amplitude as a function of cognitive activity. Perception and Psychophysics, 1973, 14. 319-324.

Donchin, E., McCarthy, G., Kutas, M., & Ritter, W. Event-related brain potentials in the study of consciousness. In R. Davidson, G. Schwarta, & D. Shapiro (Eds.), Consciousness and self regulation (Vol. 3). New York: Plenum Press, 1983.

Donchin, E., McCarthy, G., & Kutas, M. Electroencephalographic investigations of hemispheric specialization. In J. E. Desmedt (Ed.), Language and hemispheric specialization in man: Cerebral ERPs. Progress in Clinical Neurophysiology (Vol. 3). Karger: Basel, 1977.

Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), Brain event-related potentials in man. New York: Academic Press, 1978.

Donders, F. C. On the speed of mental processes. W. G. Kosler (Trans.) Acta Psychologica, 1969, 30, 412-431.

Coombs, C. H., Dawes, R. M., & Tversky, A. Mathematical psychology: An elementary introduction. Englewood Cliffs, NJ: Prentice Hall, 1970.

Davis, R. The limits of the "psychological refractory period." Quarterly Journal of Experimental Psychology, 1956, 8, 24-38.

Davis, R. The human operator as a single channel information system. Quarterly Journal of Experimental Psychology, 1957, 9, 119-129.

Davis, R. The role of "attention" in the psychological period. Quarterly Journal of Experimental Psychology, 1959, 11, 211-220.

Deutsch, J. A., & Deutsch, D. Attention: Some theoretical considerations. Psychological Review, 1963, 70, 80-90.

Dixon, N. F. Preconscious processing. New York: John & Sons, 1981.

Donchin, E. On evoked potentials, cognition, and memory. Science, 1975, 190, 1004-1005.

Donchin, E. Brain electrical correlates of pattern recognition. In G. F. Inbar (Ed.), Signal analysis and pattern recognition in biomedical engineering. New York: John Wiley & Sons, 1975.

Donchin, E. Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked potentials and behavior. New York: Plenum Press, 1979.

Donchin, E. Surprise! . . . Surprise? Psychophysiology, 1981, 18, 493-513.

Donchin, E. The dissociation of electrophysiology and behavior: A disaster or a challenge? In E. Donchin (Ed.), Cognitive psychophysiology: Event-related potentials and the study of cognition. The Carmel Conferences (Vol. 1). Hillsdale, NJ: Erlbaum, 1984. (a)

Broadbent, D. E., & Gregory, M. On the interaction of S-R compatibility with other variables affecting reaction time. British Journal of Psychology, 1965, 56, 61-67.

Brooks, L. R. The suppression of visualization by reading. Quarterly Journal of Experimental Psychology, 1967, 19, 289-299.

Brooks, L. R. Spatial and verbal components of the act of recall. Canadian Journal of Psychology/Review of Canadian Psychology, 1968, 22, 349-368.

Cannon, W. B. The wisdom of the body. New York: W. W. Norton, 1932.

Carr, T. H. Consciousness in models of human information processing: Primary memory executive control and input regulation. In G. Underwood & R. Stevens (Eds.), Aspects of consciousness, Vol. 1, Psychological Issues. New York: Academic Press, 1979.

Casali, J. G., Wierwille, W. W., & Cordes, R. E. Respiratory measurement: Overview and new instrumentation. Behavior Research and Instrumentation, 1983, 15, 401-405.

Casali, J. G., & Wierwille, W. W. A sensitivity intrusion comparison of mental workload estimation techniques using flight task emphasizing perceptual piloting activities. In Proceedings of the 1982 International IEEE Conference on Cybernetics. New York: IEEE, 1982, 598-602.

Cherry, E. C. On human communication: A review, a summary, and a criticism. Cambridge: MIT Technology Press. New York: John Wiley & Sons, 1957.

Conrad, R. Speed stress. In W. F. Floyd & A. T. Welford (Eds.), Symposium on Human Factors in Equipment Design. London: H. K. Lewis & Co. for the Ergonomic Research Society, 1954.

Berlyne, D. E. Attention as a problem in behavior theory. In D. E. Mostofsky (Ed.), Attention: Contemporary theory and an analysis. New York: Appleton-Century-Crofts, 1970.

Bertelson, P. S-R relationships and reaction time to new versus repeated signals in a serial task. Journal of Experimental Psychology, 1963, 65, 478-484.

Borg, G. Psychological and physiological studies of physical work. In W. T. Singlton & D. Whitefield (Eds.), Measurement of man at work. London: Taylor and Francis, 1971.

Borg, G. Subjective aspects of physical and mental load. Ergonomics, 1978, 21, 215-220.

Borg, G., & Noble, B. S. Perceived exertion. in J. H. Wilmore (Ed.), Exercise and sports sciences reviews, (Vol. 2). New York: Academic Press, 1974.

Bostock, H., & Jarvis, M. J. hanges in the form of the cerebral evoked response related to the speed of simple reaction time. Electroencephalography and Clinical Neurophysiology, 1970, 29, 137-145.

Bratfisch, O., Borg, G., & Dornic, S. Perceived item difficulty in three tests of intellectual performance capacity. Reports from the Institute of Applied Psychology, The University of Stockholm (Report No. 29), Stockholm, Sweden, 1972.

Broadbent, D. E. Perception and communication. London: Pergamon Press, 1958.

Broadbent, D. E. Decision and stress. London: Academic Press, 1971.

Broadbent, D. E. Task combinations and selective intake of information. Acta Psychologica, 1982, 50, 253-296.

REFERENCES

Allport, D. A., Antonis, B., & Reynolds, P. On the division of attention: A disproof of the single channel hypothesis. Journal of Experimental Psychology, 1972, 24, 225-235.

Anderson, J. R. Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum, 1981.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The psychology of learning and motivation, (Vol. 2). New York: Academic Press, 1968.

Attneave, F. Application of information theory to psychology. New York: Holt, 1959.

Baddeley, A. D., & Liberman, K. Spatial working memory. In R. S. Nickerson (Ed.), Attention and performance, (VIII). Hillsdale, NJ: Erlbaum, 1980.

Baty, D. Human transformation rates in one to four axis tracking. Proceedings of the 7th Annual NASA Conference on Manual Control, (NASA, SP A281), 1971.

Beatty, J. Pupillometric methods of workload evaluation: Present status and future possibilities. In R. Auffret (Ed.), Survey of methods to assess workload. (Agard Proceedings 246), London: Harford House, 1979, pp. 103-110.

Beatty, J. Task evoked pupillary responses processing load and the structure of processing resources. Psychological Bulletin, 1982, 91(2), 276-292.

Berlyne, D. E. Attention and change. British Journal of Psychology, 1951, 42, 269-278.

Berlyne, D. E. Conflict, arousal and curiosity. New York: McGraw-Hill, 1960.

Mane, A. M., Coles, M. G. H., Wickens, C. D., & Donchin, E. The use of the additive factors methodology in the analysis of skill. Proceedings of the 27th Annual Meeting of the Human Factors Society, 1983, 407-411.

Marcel, G. Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In R. S. Nickerson (Ed.), Attention and performance VIII. Hillsdale, NJ: Erlbaum, 1980.

Martin, D. W. Residual processing capacity during verbal organization in memory. Journal of Verbal Learning and Verbal Behavior, 1970, 9. 391-397.

McCarthy, G., & Donchin, E. A metric for thought: A comparison of P300 latency and reaction time. Science, 1981, 211, 77-80.

McClelland, J. L. On the time relations and mental processes: An examination of systems of processes in cascade. Psychological Review, 1979, 86, 287-330.

McDonald, L. B. A model for predicting driver workload in the freeway environment: A feasibility study. Unpublished doctoral dissertation, Texas A & M University, 1973.

McDonnell, J. D. Pilot rating techniques for the estimation and evaluation of landing qualities. (AFFDL-TR-68-76), Wright-Patterson Air Force Base, Ohio, Air Force Flight Dynamics Laboratory, 1968.

McLeod, P. A dual task response modality effect: Support for multi-processor models of attention. Quarterly Journal of Experimental Psychology, 1972, 29, 651-667.

McLeod, P. Does probing RT measure central procewsing demand? Quarterly Journal of Experimental Psychology, 1978, 30, 83-89.

Michon, J. A. A note on the measurement of perceptual motor load. Ergonomics, 1964, 7, 461-463.

Michon, J. A. Tapping regularity as a measure of perceptual motor load. Ergonomics, 1966, 9, 401-412.

Michon, J. A., & von Doore, H. Equipment note: A semi-portable apparatus for the measurement of perceptual motor load. Ergonomics, 1967, 10, 67-72.

Miller, G. A.The magical number seven, plus or minus two. Psychological Review, 1956, 63, 87-97.

Moray, N. Attention in dichotic listening: Affective cues and the influence of instruction. Quarterly Journal of Experimental Psychology, 1959, 11, 56-60.

Moray, N. Where is capacity limited? A survey and a model. Acta Psychologica, 1967, 27, 84-92.

Moray, N. Mental workload: Its theory and application. New York: Plenum Press, 1979.

Moray, N. Subjective mental workload. Human Factors, 1982, 24, 25-40.

Moray, N., Johansen, J., Pew, R. W., Rasmussen, J., Sanders, A. F., & Wickens, C. D. Report of the experimental psychology group. In N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum Press, 1979.

Moscovitz, M., & Klein, D. Material specific perceptual interference for visual words and faces: Implications for models of capacity limitations attention and laterality. Journal of Experimental Psychology: Human Perception and Performance, 1980, 6, 590-604.

Mulder, G. Sinusarthymia and mental workload. In N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum Press, 1979, 327-343.

Mulder, G., & Mulder, L. J. M. Information processing and cardiovascular control. Psychophysiology, 1981, 18, 392-402.

Murrell, K. F. H. Ergonomics. London: Chapman & Hall, 1969.

Navon, D., & Gopher D. On the economy of the human processing system. Psychological Review, 1979, 86, 214-255.

Navon, D., & Gopher, D. Task difficulty resources and dual task performance. In R. S. Nickerson (Ed.), Attention and performance, VIII. Hillsdale, NJ: Erlbaum Associates, 1980.

Navon, D., Gopher, D., Spitz, G., & Chillag, N. On separability and interference between tracking dimensions in dual-axis tracking. Journal of Motor Behavior, in press.

Neisser, V. Cognition and reality: Principles and implications of cognitive psychology. San Francisco: Freeman, 1976.

Nisbett, R. E. & Willson, T. D. Telling more than we can know: Valid reports on mental processes. Psychological Review, 1977, 84, 231-259.

Norman, D. A. Toward a theory of memory and attention. Psychological Review, 1968, 75, 522-536.

Norman, D. A., & Bobrow, D. J. On a data limited and resources limited processes. Cognitive Psychology, 1975, 7, 44-64.

Norman, D. A., & Shallice, T. Attention to action: Willed and automatic control of behavior. In M. Lansman & E. Hunt (Eds.), Proceedings of the Lake Wilderness Attention Conference, July, 1981.

Ogden, G. D., Levine, J. M., & Eisner, E. J. Measurement of workload by secondary task. Human Factors, 1979, 5, 529-548.

Pew, R. W. Human perceptual motor performance. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. New York: Erlbaum, 1979.

Picton, T. W., & Stuss, P. T. The component structore of the human event-related potentials. In H. H. Kornhuber & L. Deecke (Eds.), Motivation, motor and sensory processes of the brain: Electrical potentials, behavioral and clinical use. Netherlands: North Holland Biomedical Press, 1980.

Pollack, I. The information of elementary auditory display I. Journal of Acoustical Society of America, 1952, 24, 745-749.

Pollack, I. The information of elementary auditors display II. Journal of Acoustical Society of America, 1953, 25, 765-769.

Pollack, I., & Ficks, L. Information of elementary multidimensional auditory display. Journal of Acoustical Society of America, 1954, 26, 155-158.

Posner, M. I. Chronometric exploration of mind. Hillsdale, NJ:Erlbaum, 1978.

Posner, M. I., & Boies, S. J. Components of attention. Psychological Review, 1971, 78, 391-401.

Posner, M. I., & McLeod, P. Information processing models: In search of elementary operators. Annual Review of Psychology, 1982, 33, 477-514.

Posner, M. I., & Snyder, C. R. Attention and cognitive control. In R. L. Solso (Ed.), Information processing and cognition. Hillsdale, NJ: Erlbaum, 1975.

Pribram, K. H., & McGuinness, D. Arousal activation and effort in the control of attention. Psychological Review, 1975, 82, 116-149.

Regan, D. Evoked potentials in psychology, sensory physiology, and clinical medicine. London: Chapman and Hall, 1972.

Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T. Application of conjoint measurement to workload scale development. Proceedings of the Human Factors Society Annual Meeting, 1981, 522-526.

Reisberg, D. General mental resources and perceptual judgment. Journal of Experimental Psychology: Human Perception and Performance, 1983, 9, 966-979.

Reisberg, D., Rappaport, L., O'Shaughnessy, M. The limits of working memory: The digit span. Journal of Experimental Psychology: Learning and Memory, in press.

Ritter, W., Simson, R., Vaughan, H. G. Association cortex potentials and reaction time in auditory discriminations. Electroencephalography and Clinical Neurophysiology, 1972, 33, 547-557.

Rohrbaugh, J. W., Donchin, E., & Ericksen, C. W. Decision making and the P300 component of the cortical evoked response. Perception and Psychophysics, 1974, 15, 368-374.

Rolfe, J. M. The secondary task as a measure of mental load in W. T. Singleton, J. A. Fox, and D. Whitfield (Eds.), Measurement of man at work, London: Taylor & Francis, 1973.

Sanders, A. F. Some variables affecting the relation between relative stimulus frequency and choice reaction time. In A. F. Sanders (Ed.), Attention and performance III. Amsterdam: North Holland, 1970.

Sanders, A. F. Some remarks on mental load. In N. Moray (Ed.), Mental workload. New York: Plenum Press, 1979.

Sanders, A. F. Stage analysis of reaction processes. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland, 1980.

Sanders, A. F. Stress and human performance, in G. Salrendy & E. Smith (Eds.), Machine pacing and occupational stress. London: Taylor & Francis, 1981.

Sanders, A. F. Ten symposia on attention and performance; Some issues and trends. In H. Bowma & D. Bauhier (Eds.), Attention and performance X. Hillsdale, NJ: Erlbaum, 1983 (a).

Sanders, A. F. Toward a model of stress and human performance. Acta Psychologica, 1983, (b).

Sanders, A. F., Wijmen, J. L. C., & Van Arkel, A. E. An additive factors analysis of the effects of sleep loss on reaction time. Acta Psychologica, 1982, 51, 41-59.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. Automatic and control processing and attention. In R. Parasuraman, R. Davis, & J. Betty (Eds.), Varieties of attention. New York: Academic Press, 1983.

Schneider, W., & Fisk, A. D. Degree of consistent training: Improvements in search performance and automatic process development. Perception and Psychophysics, 1982, 31, 160-168.

Schneider, W., & Fisk, A. D. Attention theory and mechanisms of skilled performance. In R. A. Magill (Ed.), Memory and control of action. New York: North Holland, 1983.

Schneider, W., & Shiffrin, R. W. Controlled and automatic human information processing: Decision research and attention. Psychological Review, 1977, 84, 1-66.

Shannon, C. E., & Weaver, W. A mathematical model of communication. Urbana, Illinois: University of Illinois Press, 1949.

Sheridan, T. B. Mental workload--what is it? Why bother with it? Human Factors Society Bulletin, 1980, 23, 1-2.

Sheridan, T. B., & Ferrel, W. R. Man-machine system: Information control and decision models of human performance. Cambridge, MA: MIT Press, 1974.

Shiffrin, R. M., & Domais, S. T. The development of automatism. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum, 1981.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 119-127.

Shwartz, S. P., Pomerantz, J. R., & Egeth, H. E. State and process limitations in information processing: An additive factor analysis. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 402-410.

Slater-Hammel, A. T. Psychological refractory period in simple paired responses. Research Quarterly of American Association of Health, Physical Education, and Recreation, 1958, 29, 468-481.

Sokolov, E. W. The modeling properties of the nervous system. In I. Maltzman & K. Cole (Eds.), Handbook of contemporary Soviet psychology. New York: Basic Books, 1969.

Sperando, J. C. The regulation of working methods as a function of workload among air traffic controllers. Ergonomics, 1978, 21, 193-202.

Sperling, G., & Melchner, M. J. The attention operating characteristic: Examples from visual search. Science, 1978, 202. 315-318 (a).

Sperling, G., & Melchner, M. J. Visual search attention and the attention operating characteristics. In J. Requin (Ed.), Attention and performance VII. New York: Academic Press, 1978 (b)

Squires, N. K., Donchin, E., Squires, K. C., & Grossberg, S. Bisensory stimulation: Infering decision related processes from the P300 component. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 299-315.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. On the discovery of processing stages: Some extensions of Donders' method. Acta Psychologica, 1969, 30, 276-315 (a).

Sternberg, S. Memory scanning: Mental processes revealed by reaction time experiments. American Scientist, 1969, 57, 421-457 (b)

Sutton, S., Braren, M., Zubin, J., & John, E. R. Evoked potential correlates of stimulus uncertainty. Science, 1965, 150, 1187-1188.

Teleford, C. W. The refrectory phase of voluntary and associative response. Journal of Experimental Psychology, 1931, 14, 1-35.

Tichner, E. B. Lectures on the elementary psychology of feeling and attention. New York: MacMillan, 1908.

Treisman, A. M. Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 1960, 12, 242-248.

Treisman, A. M. Verbal cues: Language and meaning in selective attention. American Journal of Psychology, 1964, 77, 206-219.

Treisman, A. M. Monitoring and storage of irrelevant images in selective attention. Journal of Verbal Learning and Verbal Behavior, 1965, 3, 449-459.

Treisman, A. M. Strategies and models of selective attention. Psychological Review, 1969, 76, 282-299.

Treisman, A. M., & Gelade, G. L. A future integration theory of attention. Cognitive Psychology, 1980, 12, 97-136.

Trumbo, D., & Milone, F. Primary task performance as a function of encoding retention and recall in secondary task. Journal of Experimental Psychology, 1971, 91, 273-278.

Underwood, G. Memory system and conscious processes. In G. Underwood & R. Stevens (Eds.), Aspects of consciousness, Vol. 1, Psychological issues. New York: Academic Press, 1979.

Vidulich, M. K., & Wickens, C. D. Time sharing manual control and memory search: The effects of input and output modality competition priorities and control order. (Tech. Rep. EPL-81-41/ONR-81-4), University of Illinois, December 1981.

Wachtel, P. L. Conception of broad and narrow attention. Psychological Bulletin, 1967, 68, 417-429.

Welford, A. T. The "psychological refrectory period" and the timing of high speed performance: A review and theory. British Journal of Psychology, 1952, 43, 2-19.

Welford, A. T. Evidence of a single channel decision mechanism limiting performance in a serial reaction task. Quarterly Journal of Experimental Psychology, 1959, 2, 193-210.

Welford, A. T. A single channel operation in the brain. Acta Psychologica, 1967, 27, 5-22.

Whitaker, L. Dual task interference as a function of cognitive processing load. Acta Psychologica, 1979, 43, 71-84.

Wickelgren, W. Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica, 1977, 41, 67-85.

Wickens, C. D. The effects of divided attention in information processing in tracking. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 1-13.

Wickens, C. D. The structure of processing resources. In R. Nickerson & R. Pew (Eds.), Attention and performance VIII. New York: Erlbaum, 1980.

Wickens, C. D. Workload: In defense of the secondary task. Personnel Training and Selection Bulletin, 1981, 2, 119-123.

Wickens, C. D. Processing resources in attention. In R. Parasuraman, J. Beatty, & R. Davies (Eds.), Varieties of attention. New York: John Wiley & Sons, 1983.

Wickens, C. D. Engineering psychology and human performance. Columbus, OH: Charles F. Merrill Publishing Co., 1984.

Wickens, C. D., & Derrick, W. The processing demands of higher order manual control: Application of additive factors methodology. (EPL-81-1/ONR-81-1), Engineering-Psychology Research Laboratory, University of Illinois, 1981.

Wickens, C. D., Derrick, W. D., Micallizi, J., & Berringer, D. The structure of processing resources. In R. E. Corrick, E. C. Haseltine, & R. T. Durst (Eds.), Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, CA, 1980.

Wickens, C. D., Gill, R., Kramer, A., Ross, W., & Donchin, E. The cognitive demands of second order manual control: Applications of the event-related brain potential. Proceedings of the 17th Annual NASA Conference on Manual Control, NASA TM, 1981.

Wickens, C. D., Israel, J. B., & Donchin, E. The event-related cortical potential as an index of task workload. Proceedings of the 21st Annual Meeting of the Human Factors Society, Santa Monica, CA, 1977.

Wickens, C. D., & Kessel, C. Processing resource demands of failure detection in dynamic systems. Journal of Experimental Psychology: Human Perception and Performance, 1980, 6, 569-577.

Wickens, C. D., & Kessel, C. Failure detection in dynamic systems. In J. Rasmussen & W. B. Rouse (Eds.), Human detection and diagnosis of system failure. New York: Plenum Press, 1981.

Wickens, C. D., Kramer, A., Vanasse, L., & Donchin, E. Performance of concurrent tasks: A psychological analysis of reciprocity of information processing resources. Science, 1983, 221, 1080-1082.

Wickens, C. D., Mountford, S. J., & Schreiner, W. S. Time sharing efficiency: Evidence for multiple resources, task hemispheric integrity and against a general ability. Human Factors, 1982, 23, 211-229.

Wickens, C. D., & Sandry, D. L. Task hemispheric integrity in dual task performance. Acta Psychologica, 1982, 52, 227-248.

Wickens, C. D., Sandry, D. L., & Micolizzi, S. A validity of the spatial variant of the Sternberg memory search task. (Tech. Rep. EPL-81-2/ONR-81-2), Engineering-Psychology Research Laboratory, University of Illinois, June 1981.

Wickens, C. D., Sandry, D. L., & Vidulich, M. Compatibility and resource competition between modality of input, central processing, and output: Testing a model of complex task performance. Human Factors, 1983, 25, 227-248.

Wickens, C. D. & Yeh, Y. Y. The dissociation between subjective workload and performance: A multiple resource approach. Proceedings of the 27th Annual Meeting of the Human Factors Society, Norfolk, VA, 1983.

Wierwille, W. W., Gutman, J. C., Hicks, T. G., & Muto, W. H. Secondary task measurements of workload as a function of simulated vehicle dynamics and driving conditions. Human Factors, 1977, 19, 557-566.

Wierwille, W. W., Skipper, J. H., & Rieger, A. C., Decision free rotating scales for workload and estimation: Theme and variations. Proceedings of the 20th Annual Conference on Manual Control, Sunnyvale, CA, 1984.

Wilkinson, R. T., & Morlock, H. C. Auditory evoked response and reaction time. Electroencephalography and Clinical Neurophysiology, 1967, 23, 50-56.

Williges, R. C., & Wierwille, W. W. Behavioral measures of aircrew mental workload. Human Factors, 1979, 21, 549-574.

Wood, C. C., & Allison, T. Interpretation of evoked potentials: A neurophysiological perspective. Canadian Journal of Psychology, 1981, 35, 113-135.

Woody, C. D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical and Biological Engineering, 1967, 5, 539-553.

Yerkes, R. M., & Dodson, J. D. The relation of strength of stimulus to rapidity of habit formation. Journal of Comparative Neurology of Psychology, 1908, 18, 459-482.

FIGURE CAPTIONS

Figure 41.1 The amount of information that is transmitted by listeners who make absolute judgments of auditory pitch. As the amount f input information is increased by increasing from two to fourteen the umber of different pitches to be judged, the amount of transmitted information approaches as its upper limit a channel capacity of about .5 bits per judgment.

Figure 41.2 Information transmitted by unidimensional absolute judgment. Data are summarized from experiments with different stimulus odalities. Subjects were asked in each trial to identify one stimulus rom the experimental set. Performance asymptotes on all modalities at evels ranged from 2 to 3 bits (4 to 8 stimuli). (Adapted from Garner, 962)

Figure 41.3 Reaction time (RT) for one subject showing the linear ependence of RT upon information rather than the number of lternatives. Subjects were asked to produce, as fast as possible, a ifferent response to each of the stimuli in the experimental set choice reaction time). Response time to a stimulus was equally nfluenced by the number of alternatives in the set, its presentation robability, or its dependent probability, given the previously resented stimulus.

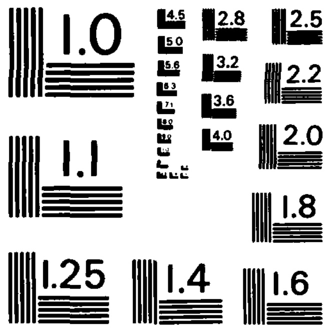
Figure 41.4 A diagram of the flow of information within the ervous system. Information received by the senses is transmitted in

AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10

5/9

UNCLASSIFIED

NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

parallel to the short-term buffer, and arrives at the filter. The filter is tuned to pass only messages having relevant physical properties, one message at a time, to the central processor. The central processor precedes the long-term store and the response mechanisms. The filter protects the central processor from overload. Screening of information is done at an early stage, before semantic analysis (the costly operation) has been performed.

Figure 41.5 Ideal plots of the latency of the second response (RT2) as a function of the interval between the first and the second stimuli (ISI), when S2 comes during the processing and response of the first stimulus (RT1). Line (a) shows the results expected when RT1 and RT2 are exactly the same in all trials. Line (b) shows the expected smoothing due to intertrial variability of RT's (after Welford, 1967). It can be seen that the predicted delay in response caused by the psychological refractory period is maximal when two stimuli are presented simultaneously but are not grouped. It diminishes in a 45-deg slope as the overlap between the two stimuli decreases. Line (c) depicts a possibility of an additional delay due to feedback processes from the first response.

Figure 41.6 Conrad's (1951) data fitted assuming that signals arrive at random intervals and that dealing with each takes an equal time t. It is also assumed that if the subject cannot deal with the signal at the instant it arrives, it can wait, but that data from not more than two signals at a time can wait in this way. In other words,

the latitude in responding is two signals. I: 2 dials, t=.37, II: 3 dials, t=.59, III: 4 dials, t=.74.

Figure 41.7 Energetical model of capacity. Performance is constrained by the availability of mental energy from a single undifferentiated pool. Note the close linkage between capacity and the physiological determinants of arousal. Note also the introduction of an allocation policy mechanism, and the closure of the loop from the present activity to the capacity pool which suggest the idea of elasticity in the total amount of available resources.

Figure 41.8 Conceptual framework of multiple resources. The overall level of competition and interference between concurrently performed tasks is determined by the degree of their overlap (sharing) on four dimensions: modality of inputs, type coding operations, stages of processing, and the nature of responses.

Figure 41.9 Cognitive-Energetical model of multiple resources. The model represents an integration of structural and energetical views. Three sources of energy supply are differentially linked to mental operations organized in independent processing stages. Arousal and activation are the main energizers of automatic processing activity. Their operation is augmented and balanced by the effort resource. Effort represents the forces of voluntary attention. It is guided by the evaluation mechanism, and is selectively energizing choice of response operations.

Figure 41.10 An example of performance based upon consistent and varied mapping. Subjects were asked to detect the presence or absence of a letter from a predesignated set in visual frames including different number of letters. Under consistent conditions, the same set of 7 letters always served as targets, and another set of 8 letters were always used to select distractors. In varied mapping, the same letters switched roles serving as targets in some blocks and distractors in others. The figure depicts mean reaction times for correct responses as a function of the number of digits on the display. Data for both negative (absence) and positive (presence) trials is presented. Compare the linear increase of response time with the increase in the number of elements in varied mapping, to the flat slopes under consistent mapping. (From Shiffrin & Schneider, 1977, experiment 2)

Figure 41.11 Three hypothetical POC curves depicting performance tradeoffs between concurrently performed tasks. Each curve traces the bounds of joint performance under different levels of intertask priorities. Combinations C1, 2, 3, 5 on and in (a) are feasible, C4 is not. Numbers in brackets indicate the priority level of each task at that point. The three curve different types of overlap between tasks in demands for processing resources. (a) Total overlap, (b) Partial overlap, (c) Complete independence.

Figure 41.12 A hypothetical example of a family of performance operating characteristics describing dual task performance of tasks X and Y with manipulation task priorities (allocation policy) and task

difficulty. Difficulty of Y is varied at three levels: Easy (E), Medium (M), and Difficult (D); task X is unchanged. Quadrate I depicts the three Performance Operating Characteristics (POCs) resulting from the combined performance of task X with the three variants of task Y. Quadrates II, III, and IV illustrate how the POCs of joint performance are related to the performance resource function of each task (Quadrate II for Task Y and Quadrate IV for X). Points A and B in quadrate III show how two different priority levels would affect joint performance.

Figure 41.13 Subject's display in concurrent performance of tracking and letter typing with manipulation of priorities. Tracking required to follow the target driven by the computer, with the X controlled by the right hand controller. Typing was performed by using the left hand keyboard to enter the motor chord codes of Hebrew letters, presented within the tracking target square. Continuous feedback on performance was given through the two horizontally moving bargraphs in the upper section of the display. A short vertical line indicated the desired level of performance and was moved to the right and left to inform subjects on changes in the emphasis on tasks. The size of the performance bargraphs reflected the instantaneous difference between actual and desired levels of performance.

Figure 41.14 An empirical example of obtained performance resource functions. Average response times for typing letters are plotted as a function of the relative priority of this task in concurrent performance with a tracking task. Priorities were changed by instructing subjects

to change their standards of performance relative to a baseline obtained for each subject. Each curve depicts the results for one of the three variants of the typing task employed in this experiment.

Figure 41.15 An experimental example of a family of performance operating characteristics. The curves depict tradeoffs under time sharing conditions with manipulation of priorities, between a constant difficulty tracking task and three variants of a letter typing task. Dotted lines were used to connect performance in dual task conditions with single task levels on the typing task.

Figure 41.16 Event Related Brain Potentials (ERPs) recorded from a group of subjects. Each trace represents the ERP obtained by averaging over trials and over subjects. These data were recorded at an electrode at the Parietal site (Pz) referred to a linked ear electrode. Positivity at the scalp electrode is reflected by a downwards deflection. Each of the superimposed pairs was recorded in experimental series in which the subject was presented with a Bernoulli sequence of tones. The solid lines in the left column were recorded when the subject was instructed to count and report, after the series was presented, the number of times the high tone was presented. This is the so called "odd-ball" task. There were 9 such series and they differed in the probability that the high tone will be presented. The probability was varied from .10 to .90 in increments of .10, as indicated by each trace (the percentage figure by each trace indicates the percentage of high tones in the series). Note that when the high

tones were rare the ERP is characterized by a large positive-going wave. This is the P300. The ERPs represented by the dotted lines superimposed on the solid lines were elicited by precisely the same tones that elicited the solid lines, except that in this case the subject was not counting tones but was rather trying to solve a word-puzzle. Note the absence of the P300 from these ERPs. These data indicate that the P300 is not an obligatory response to the tone, that its elicitation depends on the fact that the tone is task relevant and on the rareness of the tone. The data in the right column were elicited by the low tones that were presented in the series used in the corresponding pair in the left column. Note that when low tones were rare they nevertheless elicited a larger P300 even though they were not counted. (From Duncan-Johnson & Donchin, 1977.)

Figure 41.17 A schematic representation of the experimental arrangements used in studies of Workload utilizing the ERP. A primary task is controlled by the PDP 11/40 computer. In the instance shown, the subject controls a cursor's movement on a screen while the computer controls the movement of the target. As the computer monitors the subject's performance and controls the display the task can be made adaptive. That is, the difficulty of the task can be changed as the subject becomes more proficient. The Electroencephalogram is recorded and digitized in a conventional manner. Probe stimuli used in the "odd ball" task are also generated by the computer. (From Wickens, Isreal & Donchin, 1977.)

Figure 41.18 Event Related Brain Potentials (ERPs) elicited in response to probes in the arrangement shown in Figure 41.17. Subjects were tracking a target in a study in which tracking difficulty was varied in a graded fashion by varying the bandwidth of the signal that controlled the target's movement. Subjects were counting high-tones in a task similar to that described in connection with Figure 41.16. Note that when tracking is introduced the amplitude of the P300 is decreased relative to its amplitude when the subject's sole task requires counting tones. Further increases in task difficulty, however, do not further decrease P300 amplitude. (From Isreal, Chesney, Wickens & Donchin, 1980.)

Figure 41.19 These ERPs were obtained when the subject was monitoring a display in which numerous targets were moving. On occasion, one of the targets could briefly intensify. On other occasions, the target would change course. This experiment is similar to that described in Figure 41.18, except that the primary task in this case required the processing of a visual target. Task difficulty was varied by varying the number of elements in the monitored display. The secondary task assigned was again the counting of tones in an oddball paradigm. Data are shown for 8 subjects, as well as a "grand average" for all 8 subjects. Note that when the subjects are monitoring the display for course changes the amplitude of the P300 is reduced when the monitoring assignment is introduced. The amplitude is further reduced when the number of elements in the display is increased. The effect is

specific to the course-detection task. When the subject is monitoring the display for flashes the effect is much diminished. (From Isreal, Wickens, Chesney & Donchin, 1980.)

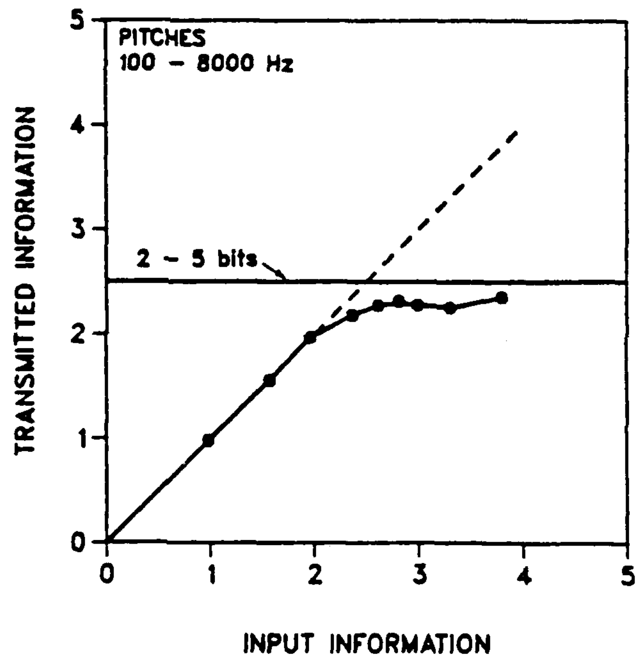
Figure 41.20 The display used by Kramer et al. Subject's task is to align the manipulator with the Target. The latter moved in a linear track. The cursor is controlled by the subject using a joy stick. When the manipulator nears the target the target begins to rotate and the subject must align his manipulator with target using a joy stick. Task difficulty could be varied by varying the order of control for the acquisition joy-stick. In this case, the secondary task utilized as probe stimuli brightenings of either the cursor or the target at different phases of the study. (From Kramer, Wickens & Donchin, 1983.)

Figure 41.21 Representative ERPs elicited in the study described in Figure 41.20. In each quadrant we are shown ERPs elicited by both target and cursor. However, in the left column are shown data recorded when the target was relevant and in the right column data recorded when the cursor was relevant. Note that in each case only the relevant stimulus elicited a P300 even though both were in the monitored part of the visual field. Note also how P300 amplitude elicited during the acquisition phase exceeds the amplitude elicited during the more demanding alignment task. Furthermore, the order of control for the joy stick also affected P300 amplitude as predicted. (From Kramer, Wickens & Donchin, 1983.)

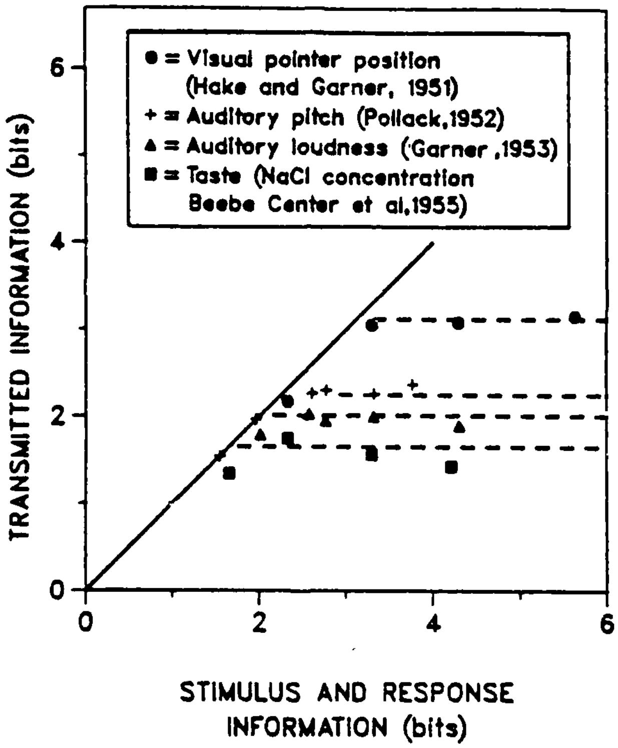
Figure 41.22 Subjects in this study performed a Step Tracking task as a "primary" task concurrently with different oddball tasks serving as "secondary" tasks. A target box made discrete jumps to the right or to the left and the subject was required to move, by means of a joy stick, a cursor onto the moving target. The difficulty of the task was varied by manipulating the relationship between the movements of the joy stick and the movement of the controlled cursor. This was combined with changes in the degree to which the movement of the target was regular, (i.e., movements in alternate directions) or random. The root mean square error, averaged over subjects, is plotted against task difficulty, as is the subject's estimate of the error involved in each of the experimental conditions. The different lines in each frame were recorded while different secondary task conditions were used. In the "auditory" and "visual" series the subjects counted either visual or auditory probes that were not associated with the step tracking. In the "counted steps" condition the subject counted the number of times the step was made in one of the two directions. Thus, in this condition the oddball stimuli were embedded in the primary task. It was predicted that in this case the amplitude of the P300 associated with the secondary task will increase with increases in primary task difficulty. (From Kramer, Wickens, Vanasse, Heffley & Donchin, 1981.)

Figure 41.23 The ERPs acquired in the Step Tracking task described in Figure 41.22. Each frame presents the ERPs elicited by secondary task probes, as indicated, for the four different levels of difficulty

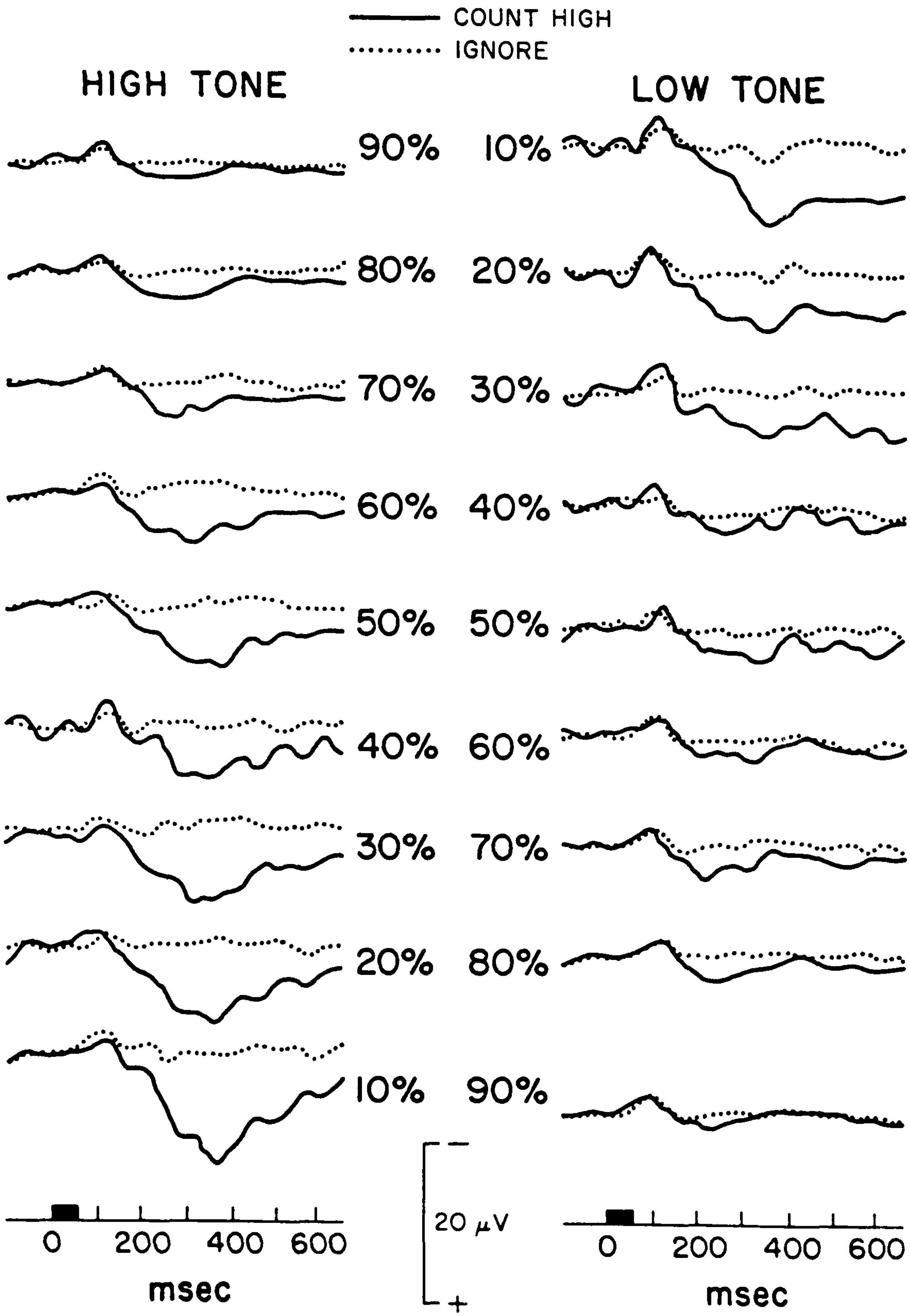
of the primary task (one being - total absence of the primary task). Note that for the Auditory and Visual conditions the data replicate previous results and P300 amplitude decreases with increasing tracking difficulty. This pattern reverses when the probes are embedded in the primary task, as they do in the "step count" condition. (From Kramer, Wickens, Vanasse, Heffley & Donchin, 1981.)



UD 1991
GOPHER
FIG NO 00.1

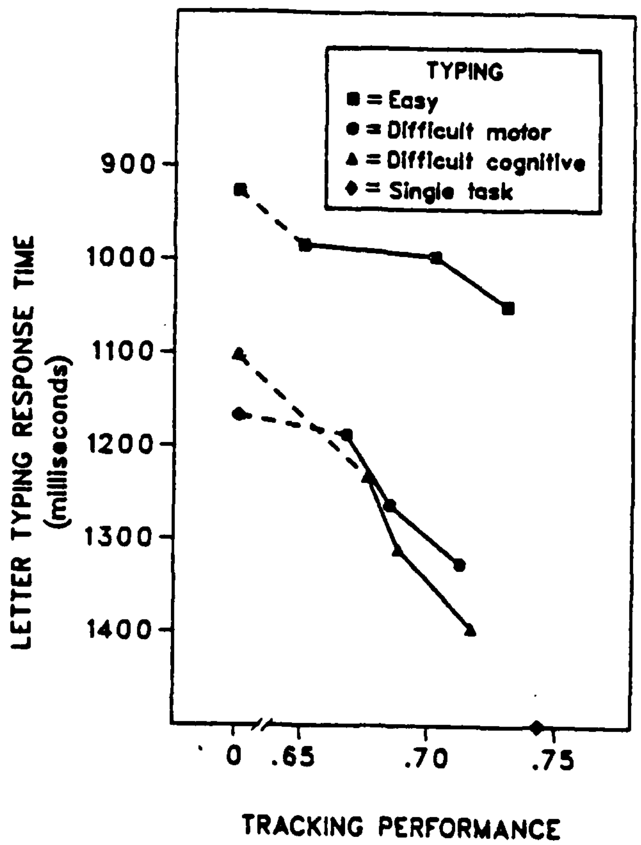


~~FEATURE~~ UD 1992
GOPHER
FIG. NO. 00.2

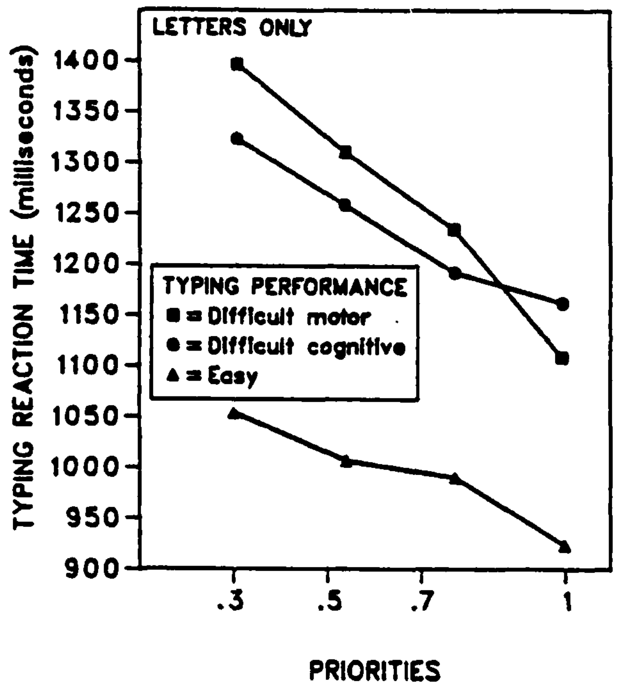


ıncan-Johnson & Donchin (1977)

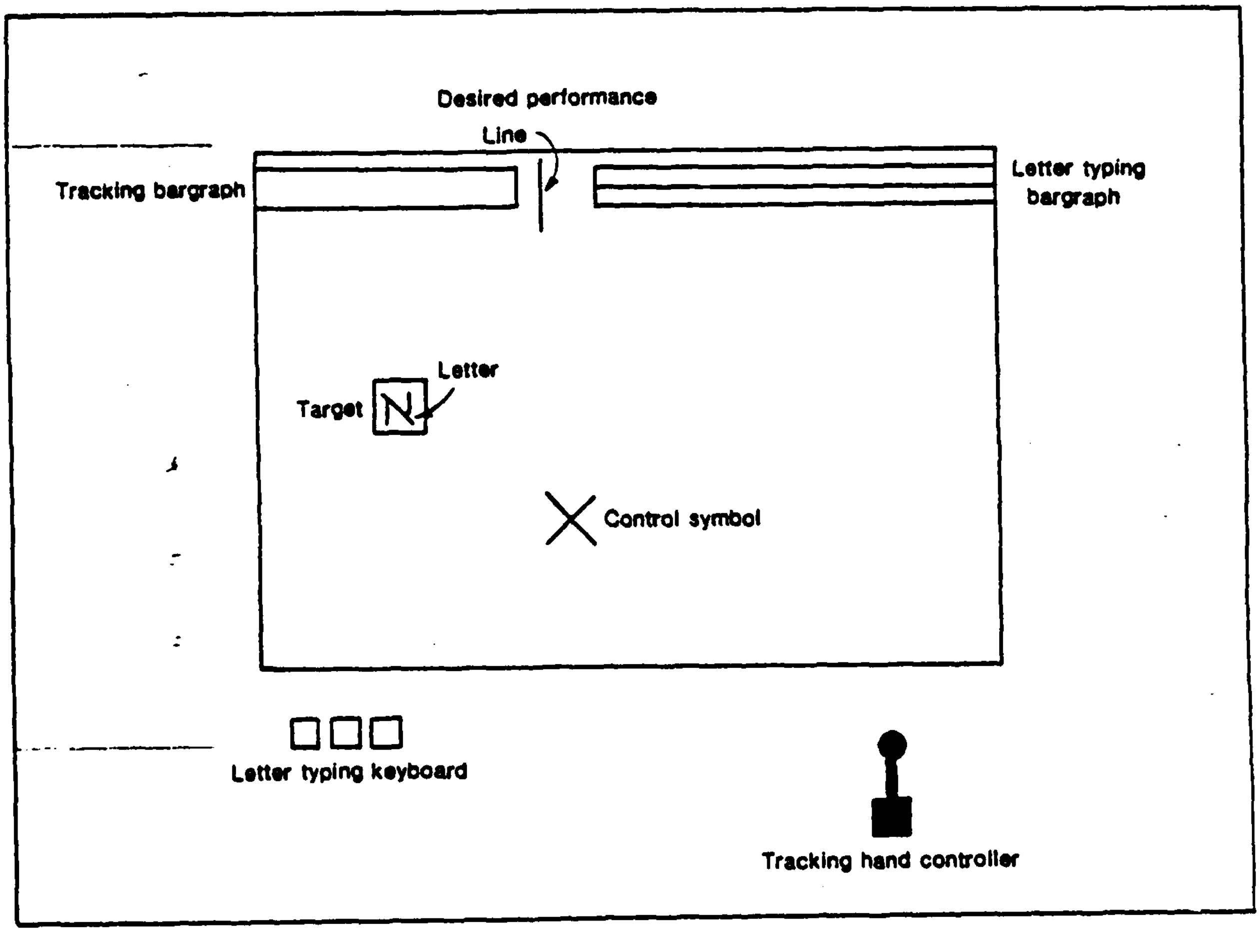
Figure 16



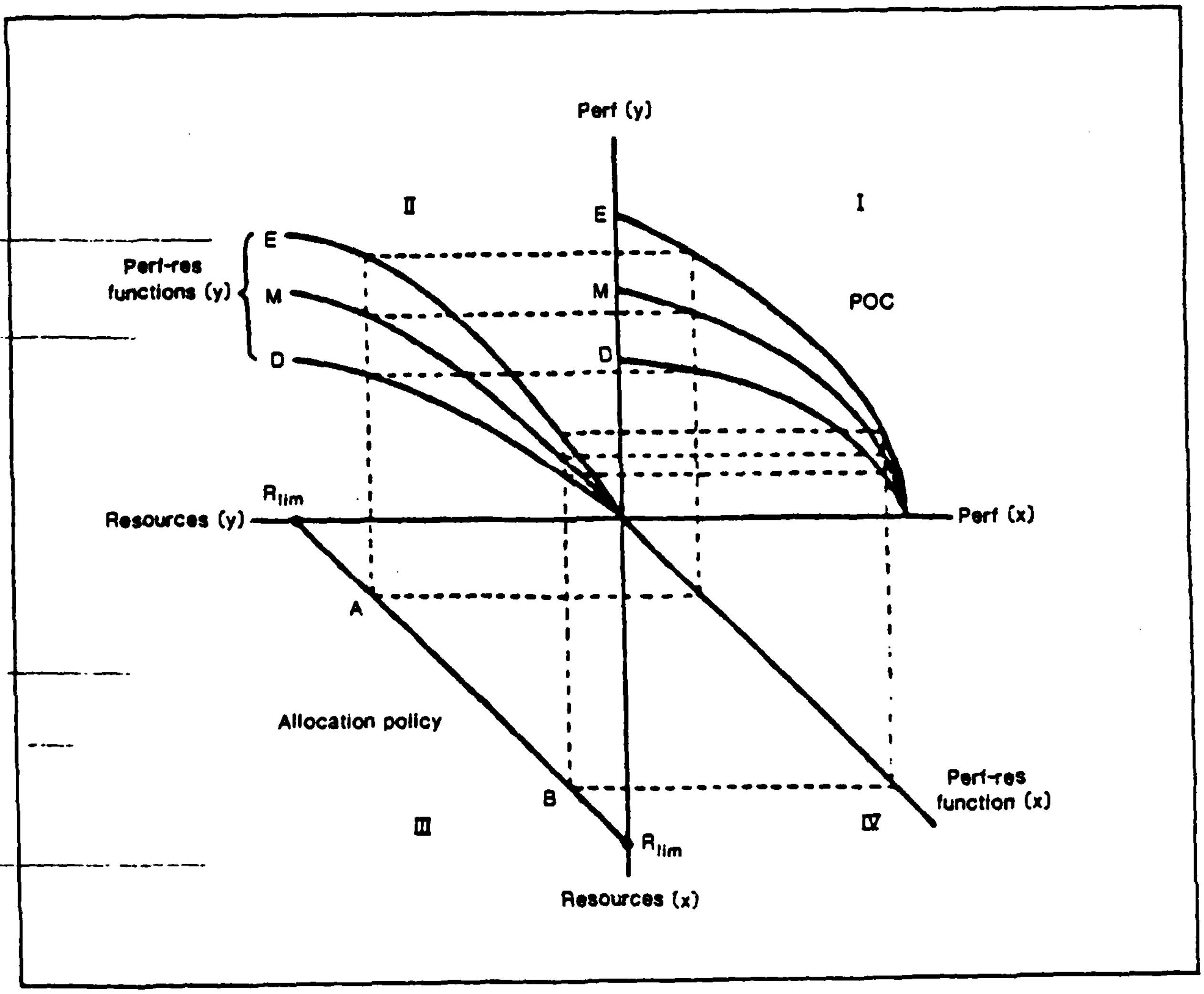
4 D 2005
GOPHER
FIG NO 00.15



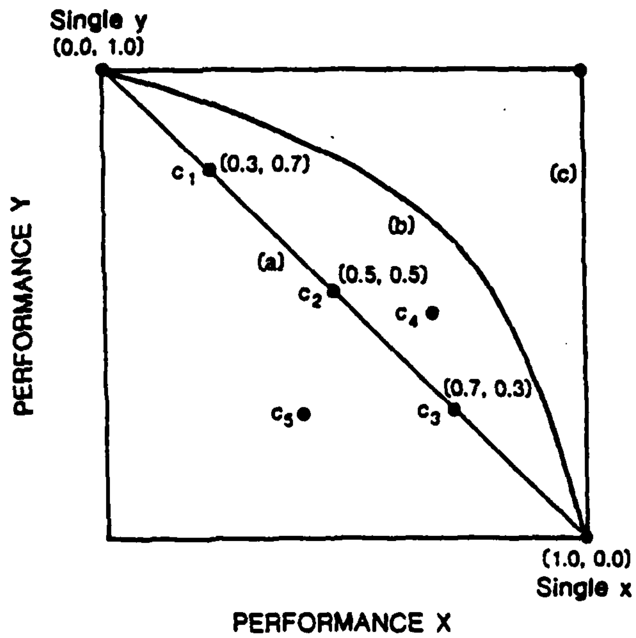
UD - 2004
GOPHER
FIG NO 00.14



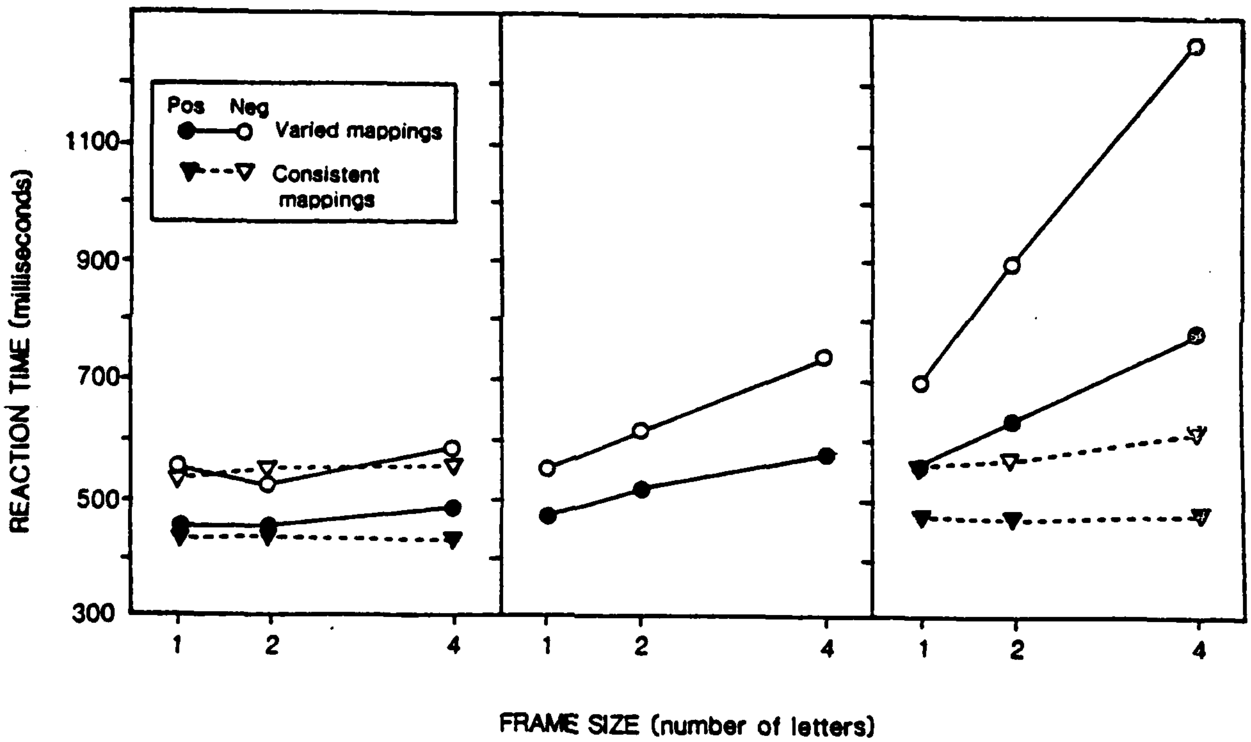
4D 2003
GOPHER
FIG NO 00.13



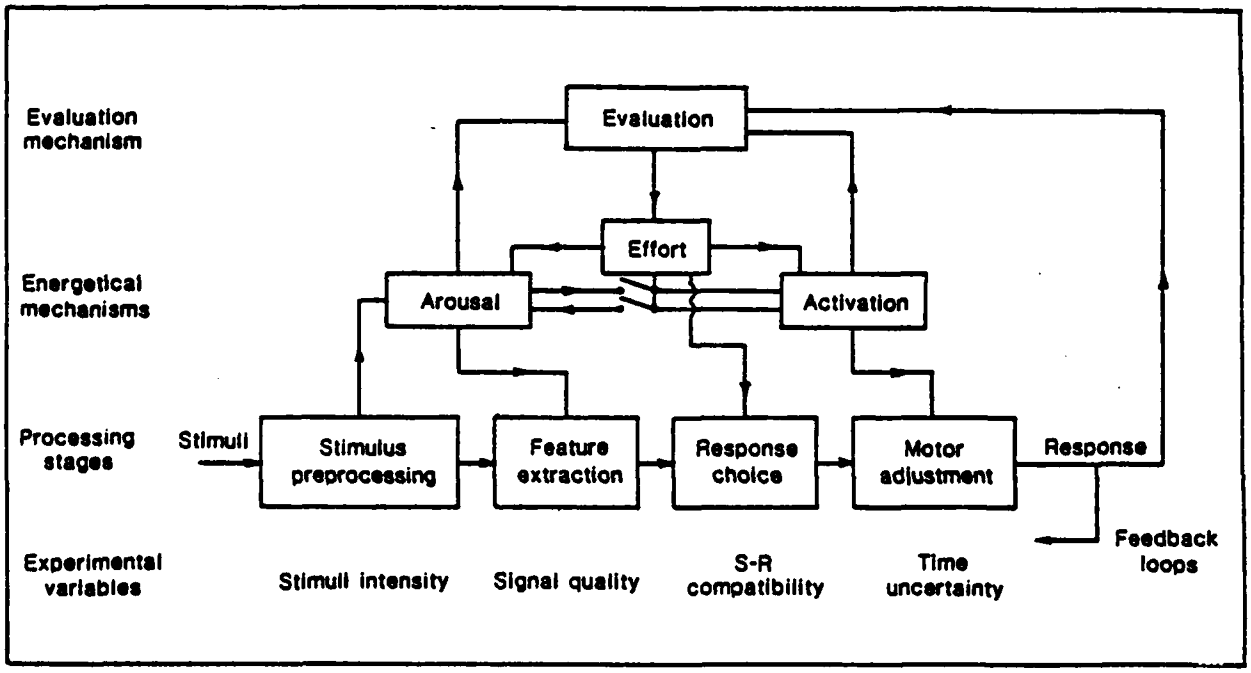
U. D. 2002
Donchin
Fig. No. 00.12



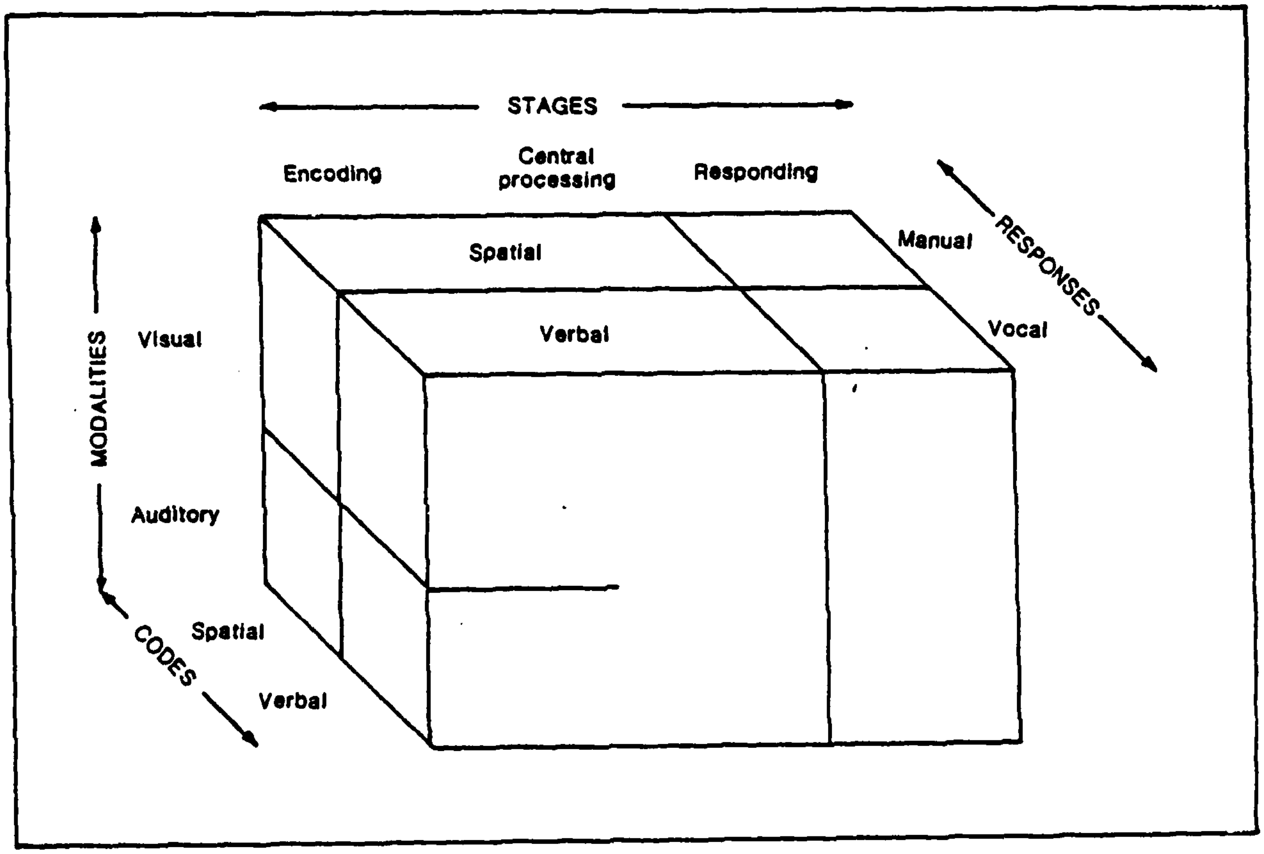
UD 2001
GOPHER



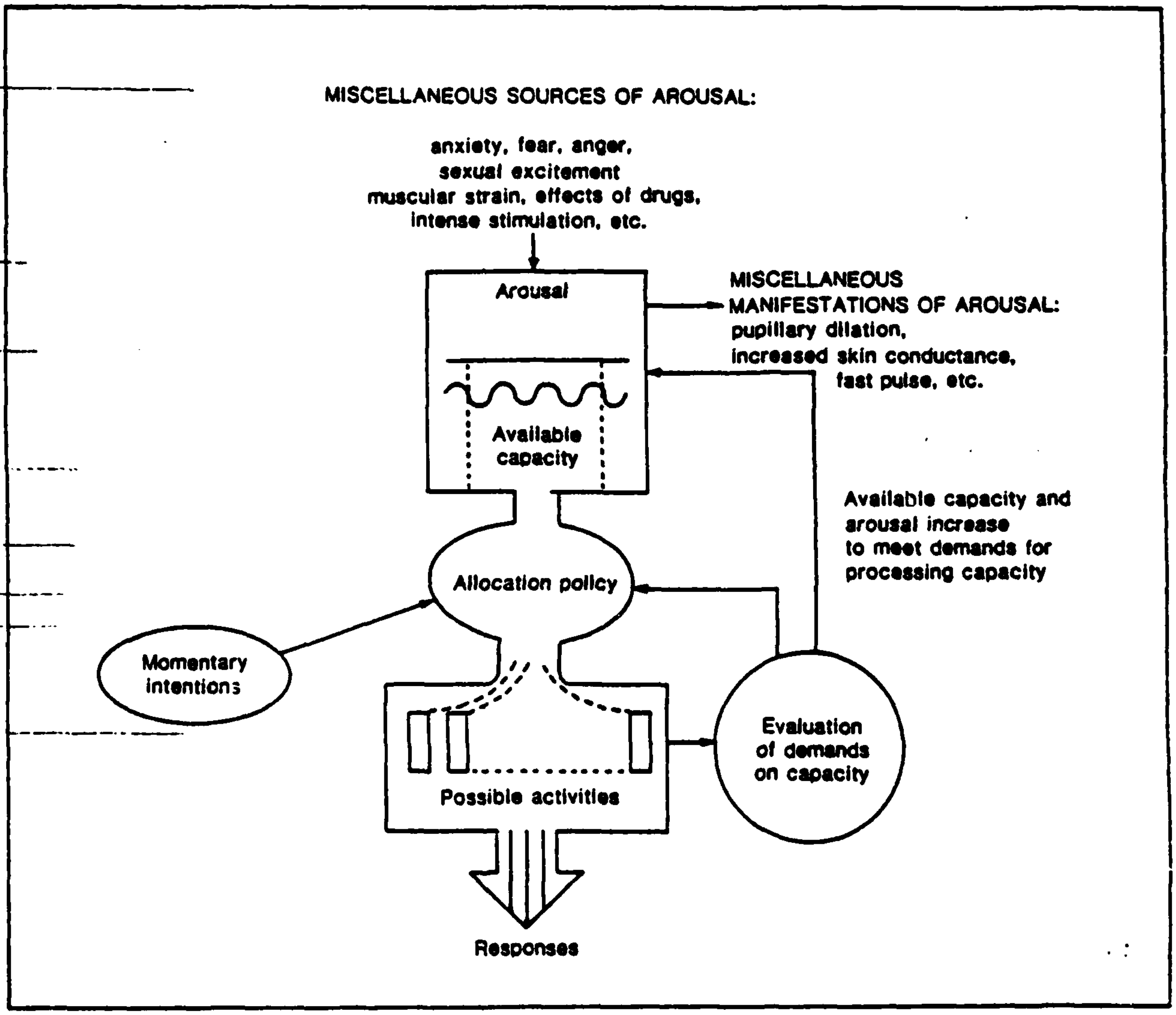
UD · 2000
GOPHER
FIG NO 00.10



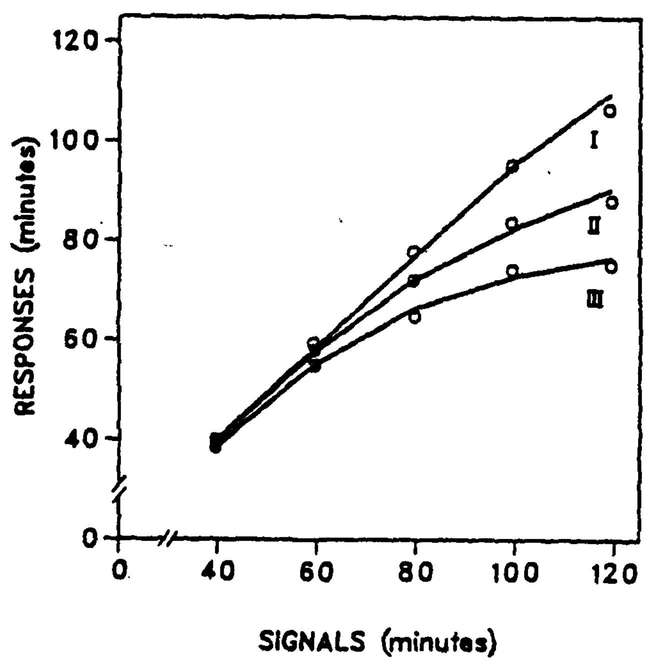
UD 1999
GOPHER
FIG. NO. 00.9



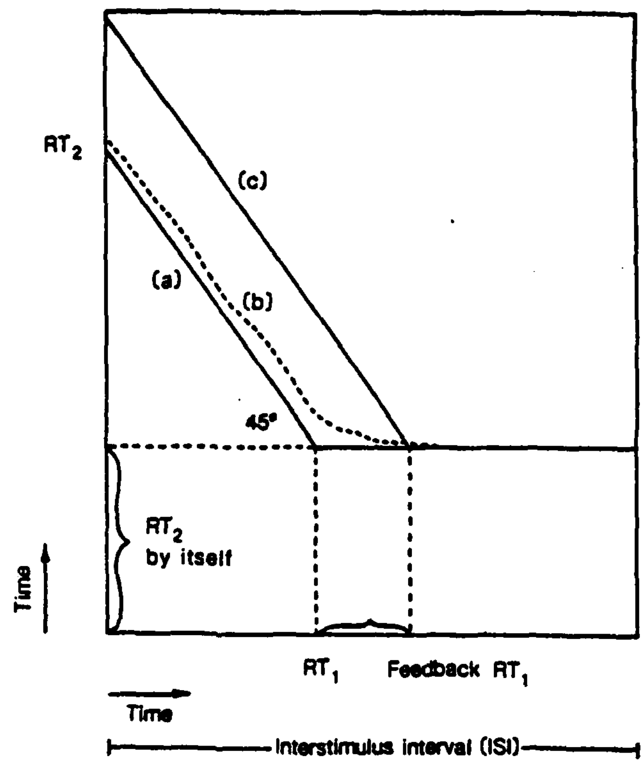
UD 1998
GOPHER



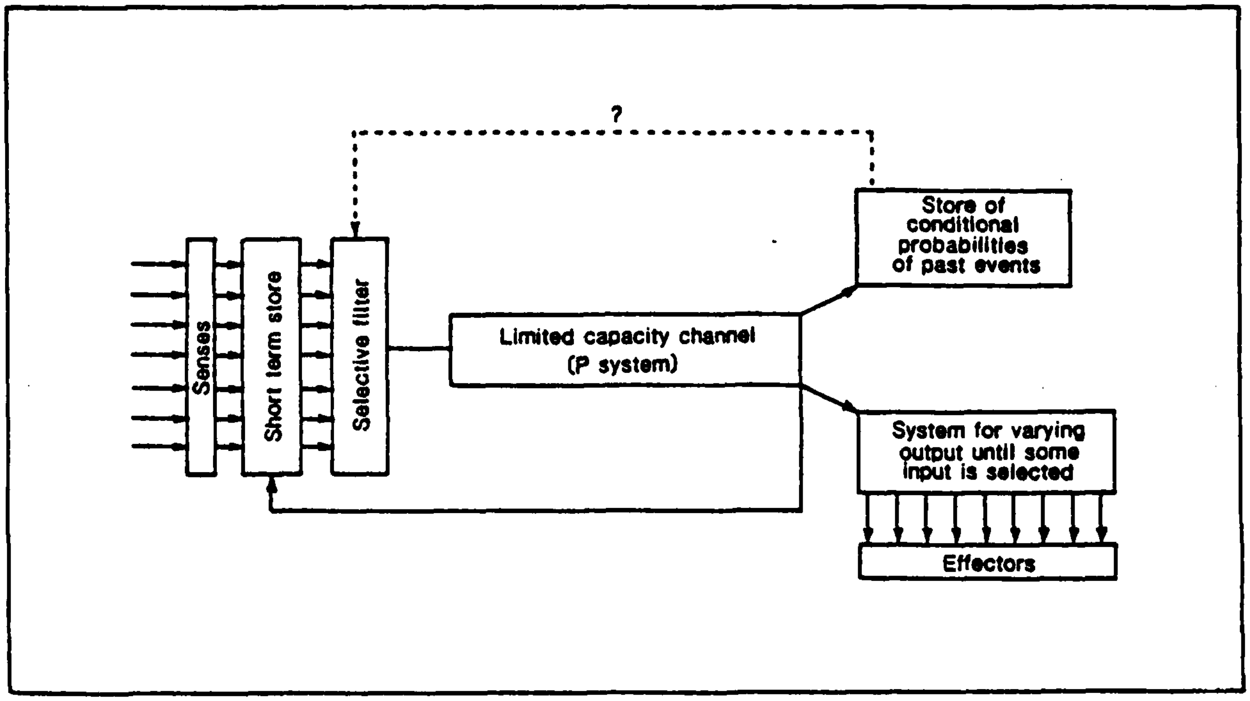
UD 1997
GOPHER
FIG. NO. 00.7



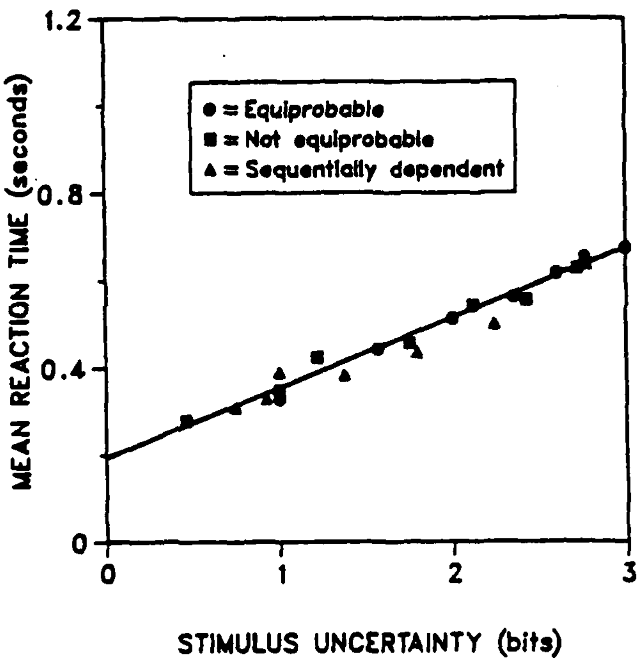
UD 1996
GOPHER
FIG. NO 00.6



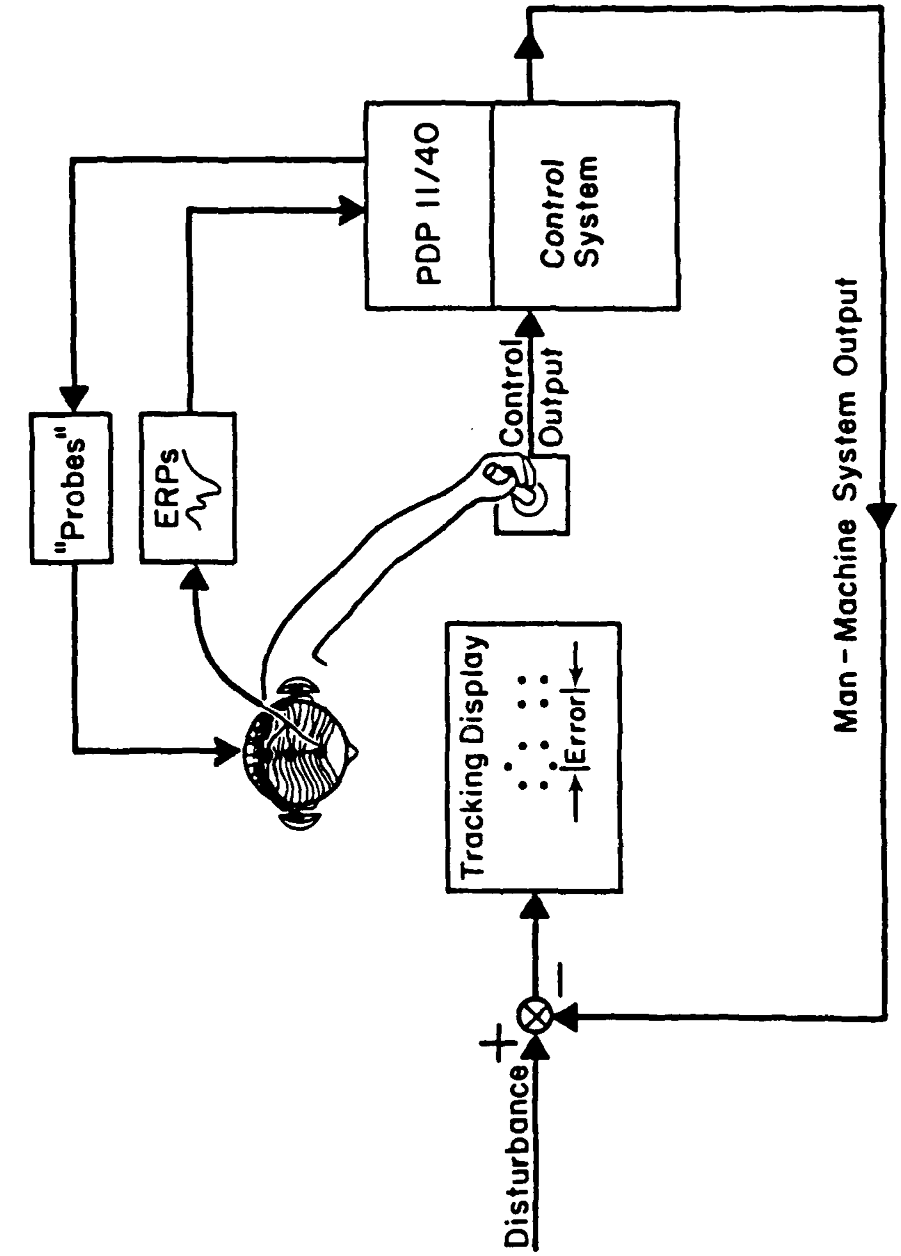
UD - 1995
GOPHER
FIG. NO. 00.5



UD 1994
GOPHER
FIG. NO. 00.4

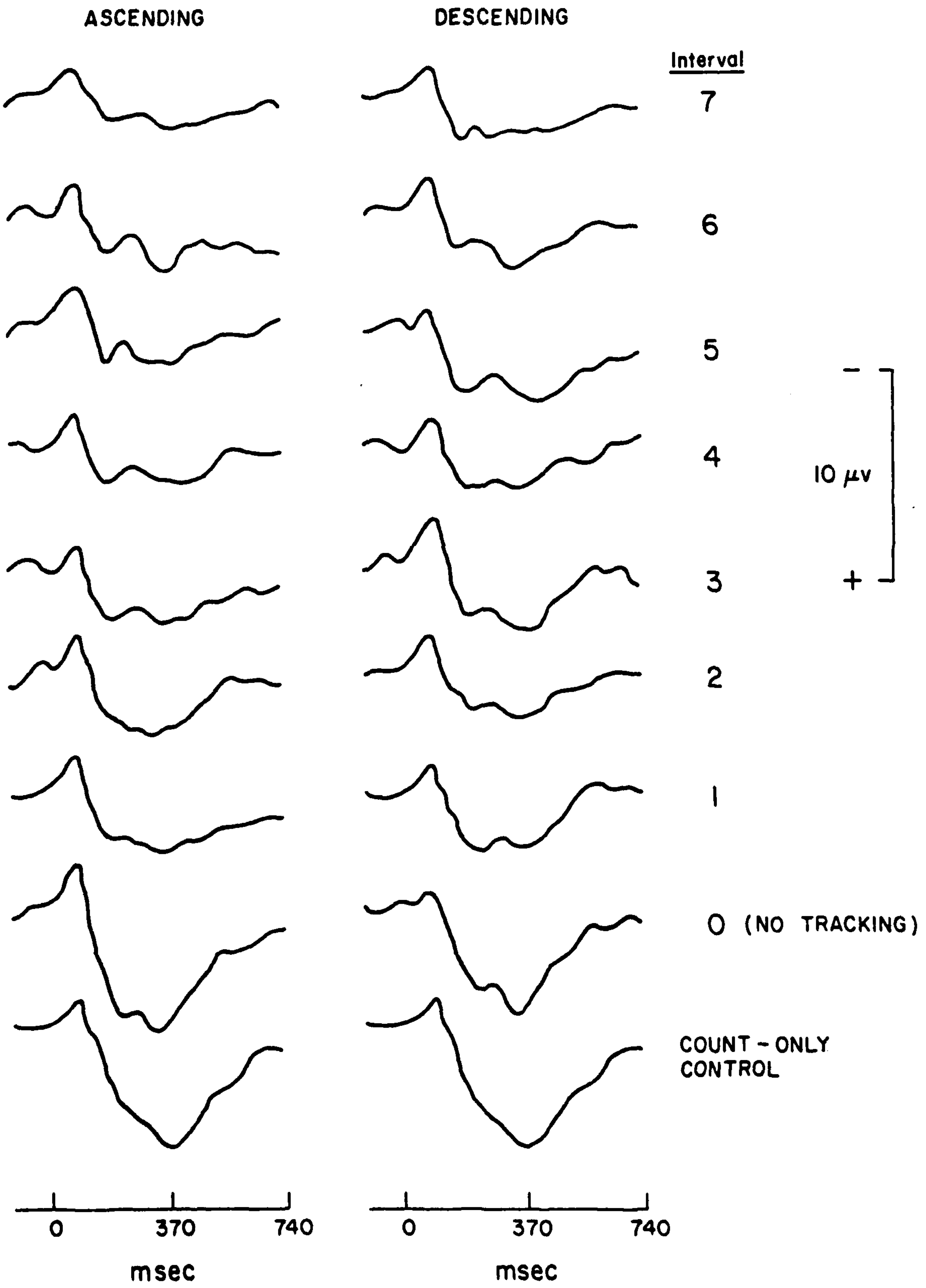


UD 1993
GOPHER
FIG. NO. 00.3



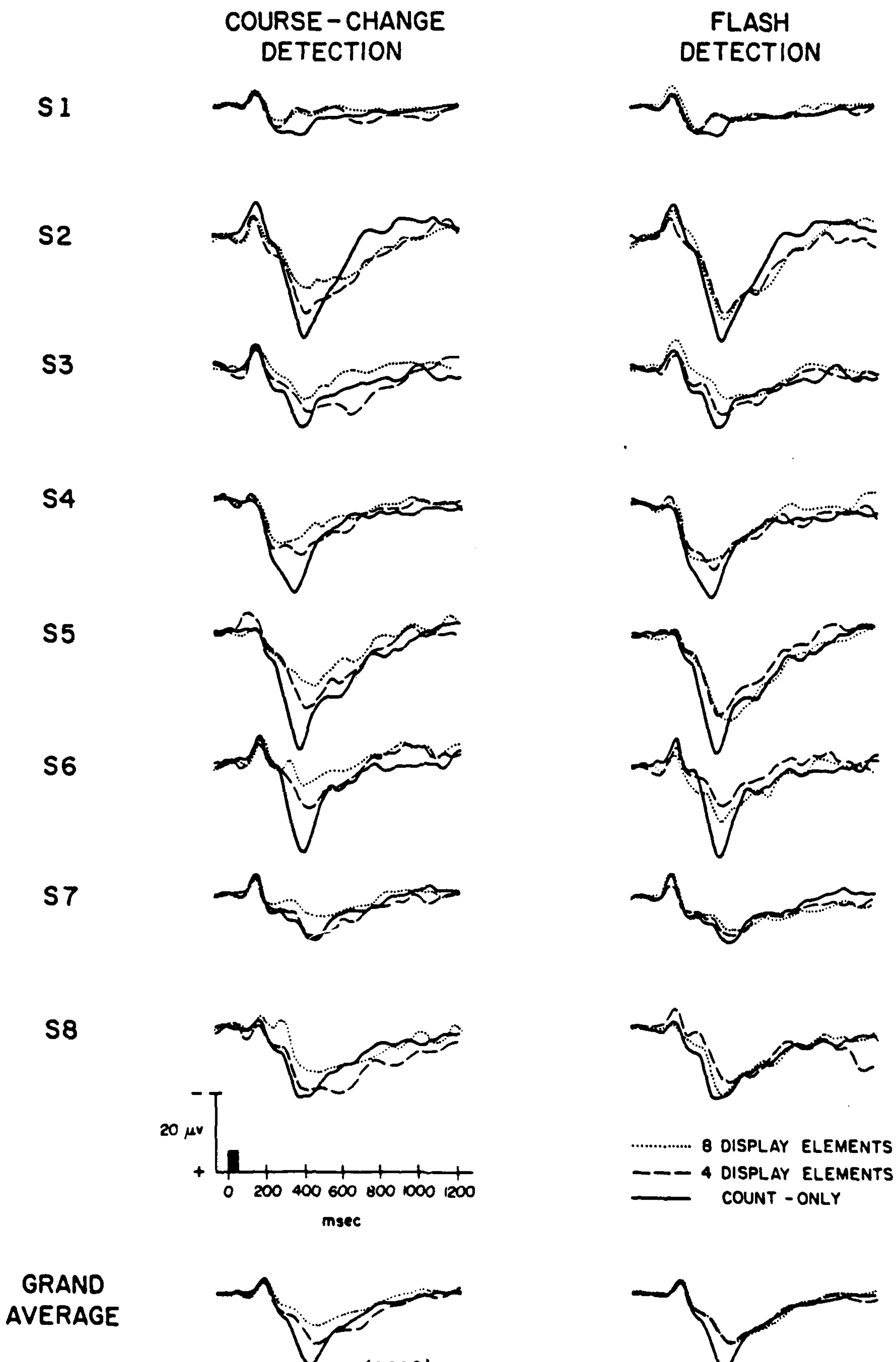
Wickens, Isreal & Donchin (1977)

Figure 17



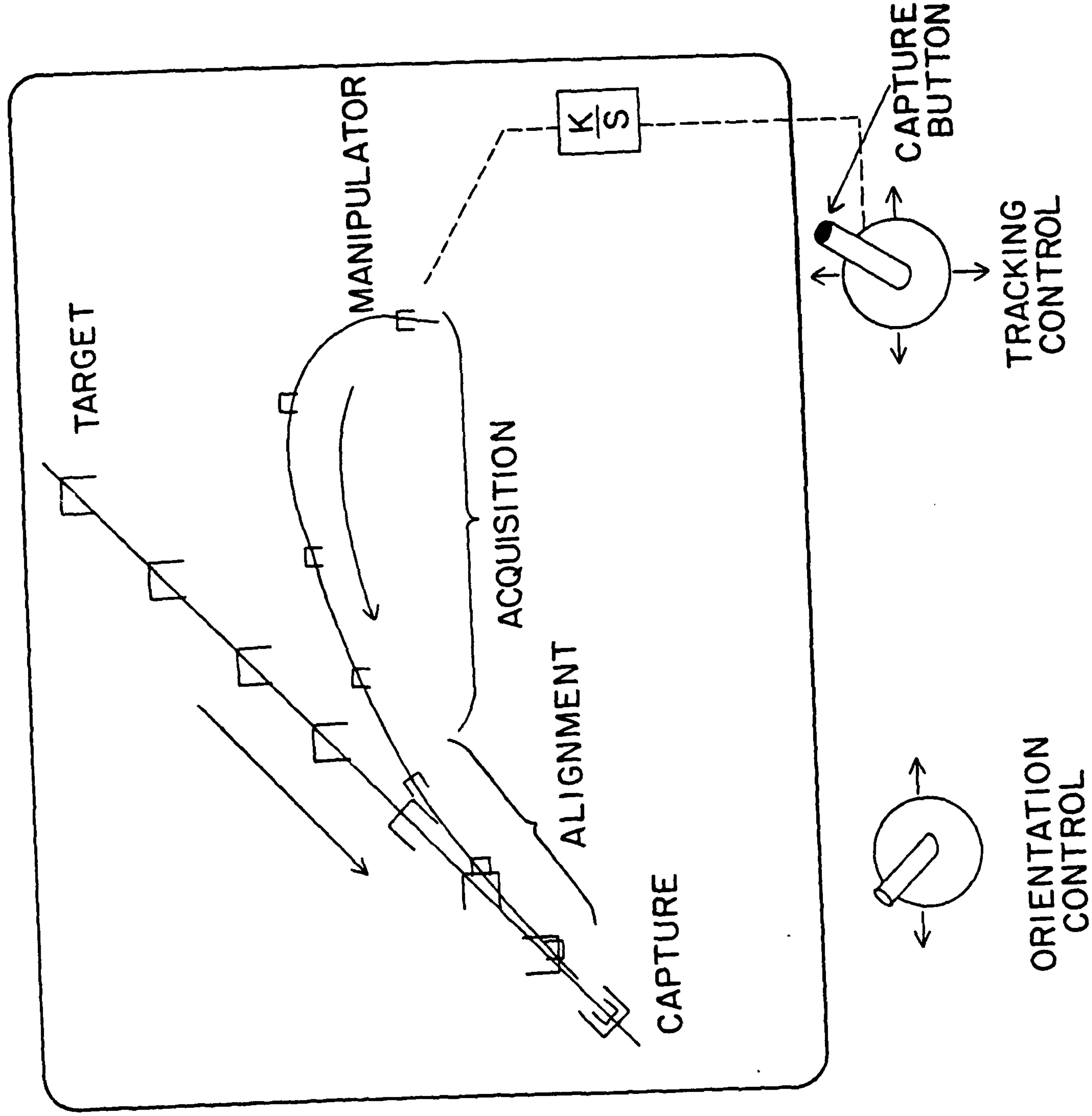
Isreal, Chesney, Wickens, & Donchin (1980)

Figure 18



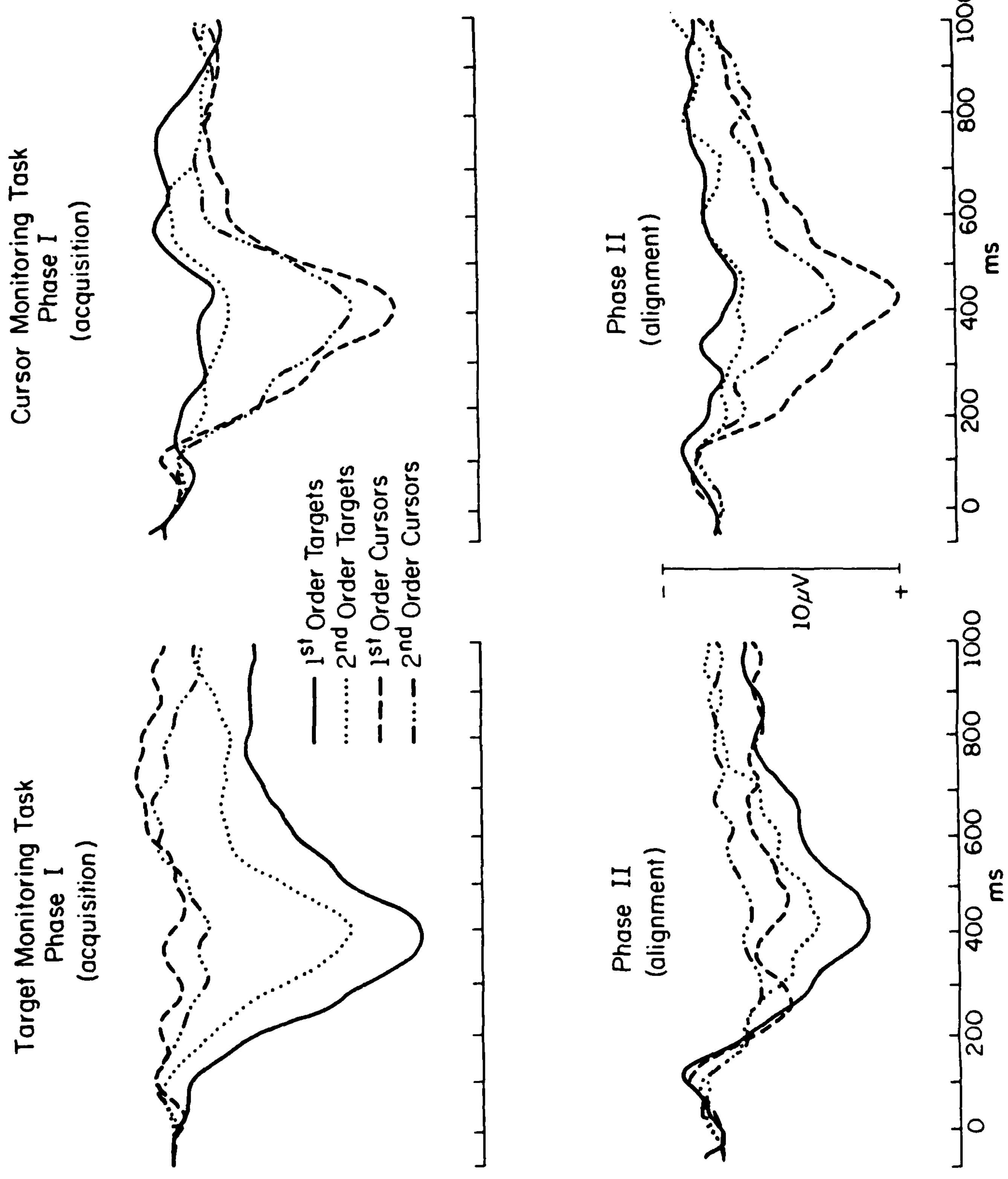
Isreal, Wickens, Chesney & Donchin (1980)

Figure 19



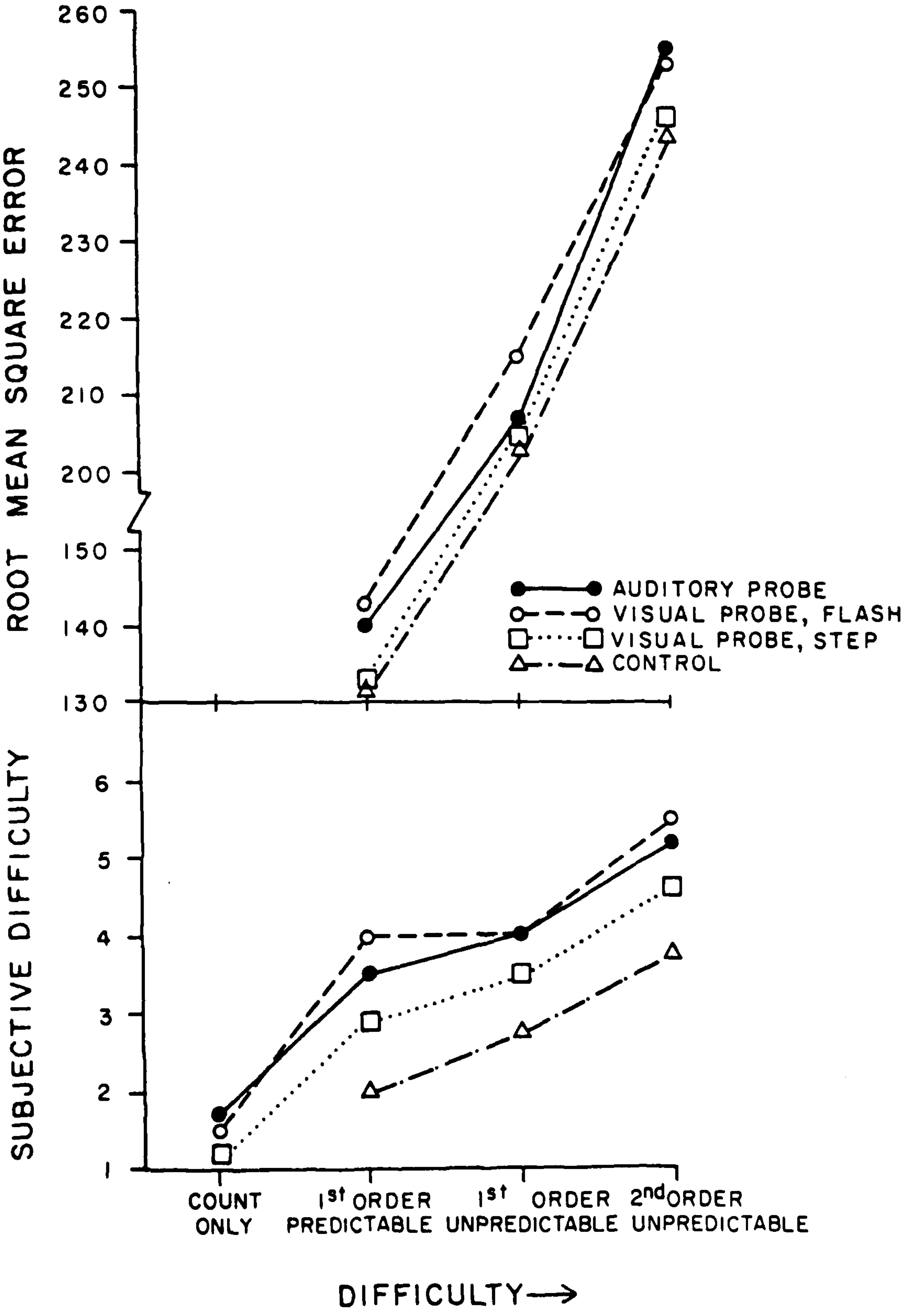
Kramer, Wickens, & Donchin (1983)

Figure 20



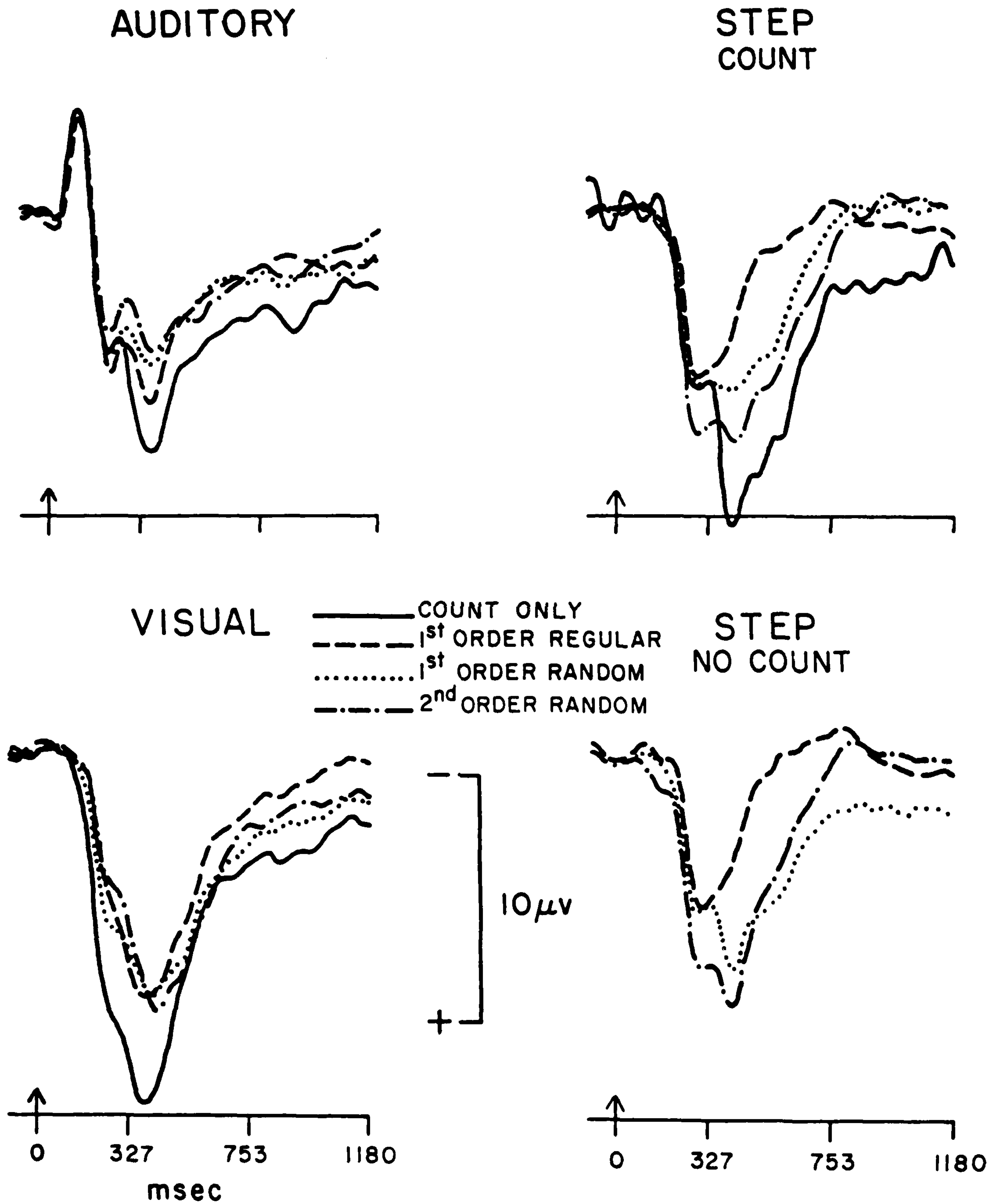
Kramer, Wickens, & Donchin (1983)

Figure 21



Kramer, Wickens, Vanasse, Heffley, & Donchin (1981)

Figure 22



Kramer, Wickens, Vanasse, Heffley, & Donchin (1981)

Figure 23

Submitted for publication. Please do not reference without permission.

A Psychophysiological Investigation of the Continuous Flow Model of Human Information Processing

Michael G. H. Coles, Gabriele Gratton, Theodore R. Bashore,
Charles W. Eriksen, and Emanuel Donchin

University of Illinois at Urbana-Champaign

Running head: PSYCHOPHYSIOLOGY AND THE CONTINUOUS FLOW MODEL

Abstract

The present experiment tested the continuous flow model of information processing by using the P300 component of the event-related brain potential to assess the duration of stimulus evaluation processes, and measures of the electromyogram (EMG) and response force to decompose response processes.

Subjects were required to respond to target letters "H" or "S" by squeezing dynamometers with the left or right hand. Target letters could be surrounded by compatible (e.g. HHHHH) or incompatible noise (SSHSS) letters. A warning tone preceded presentation of the letter array by 1000 ms on half the trials. For each trial, latency measures were available for the P300, and correct and incorrect EMG and squeeze activity. The latter measures were also used to classify each trial according to a "degree of error" dimension.

When incorrect squeeze activity was present, execution of the correct response was prolonged, indicating a process of response competition. This process occurred more often under incompatible conditions, which were also associated with a delayed P300. Thus, the noise/compatibility manipulation influenced both stimulus evaluation and response competition processes. In contrast, the warning tone increased response speed without influencing evaluation time.

These data are consistent with the continuous flow conception and suggest that the latency and accuracy of overt behavioral responses are a function of (a) a response activation process continuously controlled by an evaluation process that accumulates evidence gradually, (b) a response activation, or priming, process that is independent of stimulus evaluation, and (c) a response competition process.

A Psychophysiological Investigation of the Continuous Flow Model of Human Information Processing

Most attempts to model the human information processing system assume that it can be viewed as an ensemble of processors, each of which is responsible for performing some distinct function such as "feature detection", "stimulus encoding", or "response selection." However, models differ in the manner in which these elementary processors interact. An influential view that derives from Donders (1868/1969), and which has been refined and elaborated by Sternberg (1969), considers it possible to model the system as if the elementary processors operate serially. According to this view, a processing element (i.e., a stage) is activated upon the completion of processing by the preceding element. Information is transferred from one element of the series to the next in an all-or-none fashion. For this reason no two elements of the system are ever active at the same time.

An alternative model, proposed in different guises by several investigators, is based on the assumption that the processors function in parallel rather than in sequence (e.g. Eriksen & Schultz, 1979; Grice, Nullmeyer, & Spiker, 1982; Grossberg, 1982; McClelland, 1979; Turvey, 1973). Most proponents of this model argue that the processors operate on the partial output of other elements invoked earlier in processing (e.g. Eriksen & Schultz, 1979; McClelland, 1979). In this sense, there is "a continuous flow" of information between processors. Furthermore, for some theorists (e.g. Grice, Nullmeyer and Spiker, 1982), different processors can be activated simultaneously following the presentation of a stimulus event.

Serials models have been quite successful in accounting for an impressive array of data (Chase, 1984). However, some research findings are

clearly inconsistent with these models (e.g. Pachella, 1974). For example, consider the data from a visual search task discussed by Eriksen and Schultz (1979). The subjects were presented with a series of five letter arrays. In each array, the center (target) letter was 'H' or 'S', and the subjects were required to press one of two buttons depending on the target letter. The other four letters surrounding the target letter could be the same as the target letter ("compatible" noise), or they could be the letter calling for the opposite response ("incompatible" noise). In a variety of studies (see Eriksen & Schultz, 1979, for a review), Eriksen has found that reaction times (RTs) are longer when the noise is incompatible than when it is compatible. This finding has been interpreted by Eriksen and Schultz (1979) in terms of a continuous flow model. Incompatible arrays contain information calling for the incorrect response. As stimulus evaluation proceeds, and before it is completed, this incorrect information is passed on to the response activation system leading to activation of the incorrect response. Although the correct response may be given ultimately, the activation of the incorrect response will interfere with the execution of the correct response (through a process of response competition), and RT will be prolonged. Since response competition effects are not present for compatible arrays, RT will be shorter.

While the continuous flow model does seem to account for the effect of noise/compatibility, it is possible that the duration of stimulus evaluation is prolonged for incompatible arrays and it is this prolongation that increases RT. In particular, the duration of the evaluation process may be influenced by the greater "complexity" of the incompatible stimulus array. This explanation is quite consistent with the serial model and does not require the invocation of a continuous flow concept.

Several attempts have been made by Eriksen and his colleagues to discriminate between these two explanations of the noise/compatibility effect. For example, the effect of differences in stimulus complexity between compatible and incompatible arrays has been controlled by assigning each of the two responses to two different stimuli. Thus, the subject may be instructed to move a lever to the left in response to an H or C, and to the right in response to an S or K. With this arrangement, compatible arrays can be as visually complex (e.g. HCH) as incompatible arrays (e.g. KCK). The data indicate that RT is determined predominantly by the compatibility of the flanking noise and not by the visual heterogeneity of the stimulus array (Eriksen & Eriksen, 1979). Similarly, neutral noise letters in the array (letters that do not call for either of the responses) appear to induce response competition to the degree that these neutral characters share features with the target letters (Eriksen & Eriksen, 1974; Yeh & Eriksen, 1984).

Although these studies suggest that the compatibility of the flanking noise is a major determinant of response latency, other data suggest that the visual complexity of the display may be important (Grice, Canham, & Shafer, 1982; Flowers & Wilcox, 1982). As is true of many other controversies in cognitive psychology, resolution of this debate is hampered by the fact that the data base derives entirely from the observation of the final outcome (RT and response accuracy) of a very complex process. Although there has been an increased sophistication in the analysis of the timing and accuracy of overt response measures (e.g. Meyer, Yantis, Osman, & Smith, 1984), such measures are unlikely to provide a complete and unambiguous description of the multiple intervening processes that determine the final overt response outcome. The problem becomes particularly complex

in the case of parallel models. This is because the activities of the processors that intervene between stimulus and response have more degrees of freedom than can be described adequately by one or two measures.

In the particular case of the noise/compatibility effect, these arguments point to a need for measures that are differentially sensitive to stimulus evaluation and response competition. Donchin and his co-workers (Donchin, 1981; Kutas, McCarthy, & Donchin, 1977) have proposed that the analysis of mental chronometry can be augmented by the incorporation of psychophysiological measures. We report here a study in which we have used RT and accuracy measures, together with two psychophysiological measures, the electromyogram (EMG) and the event-related brain potential (ERP), to obtain a detailed description of the information processing activities invoked in the Eriksen paradigm.

The recording of the EMG activity associated with different motor responses provides useful information about early aspects of response execution. In particular, EMG measures can be used to detect "responses" which are initiated but fall short of complete execution. Furthermore, the relative timing of EMG and overt motor responses may shed some light on the response competition process. When response competition occurs the time between EMG activation and response initiation may be prolonged.

The utility of EMG measures is illustrated by the results of a preliminary investigation by O'Hara, Morris, Coles, Eriksen, & Morris (1981). They measured EMG responses as well as overt motor activity (button presses) in the Eriksen paradigm. Subjects had to respond with the thumbs of the two hands as a function of the target letter. The EMG was recorded from each forearm. Trials were sorted on the basis of the flanking noise (compatible or incompatible), and the presence or absence of EMG activity on

the incorrect side. O'Hara et al. (1981) found that incorrect EMG activity was apparent more often on incompatible trials and that incorrect activity tended to precede correct EMG activity. Further, on trials when incorrect EMG activity was present, the correct EMG and motor response latencies were delayed. These data were seen as providing evidence for response competition, and some support for the continuous flow interpretation of the noise/compatibility effect. However, even when there was no evidence of response competition (that is, no EMG activation on the incorrect side), RTs were still longer for the incompatible arrays. Thus, the noise/compatibility effect could not be attributed entirely to response competition. However, it could be that response competition effects occur more centrally and are not always detectable using EMG measures. Alternatively, the noise/compatibility manipulation could influence RT because of both response competition <u>and</u> stimulus evaluation effects.

In the present experiment, we measured the latency of the P300 component of the ERP to distinguish between the effects of noise/compatibility on response competition and stimulus evaluation. Donchin (1979) has proposed that the latency of this ERP component is sensitive to the duration of stimulus evaluation and categorization processes and is largely independent of the time required for response selection and execution. Several studies have confirmed this interpretation of the latency of the P300 (e.g. Duncan-Johnson & Donchin, 1982; Kutas, McCarthy, & Donchin, 1977; Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981). Note that the P300 is not identified with the stimulus evaluation process itself. Rather we propose that the function of which P300 is believed to be a manifestation can only occur after stimulus evaluation processes are completed (Donchin, 1981; Karis, Fabiani, &

compatible noise trials. This observation was confirmed by an analysis of simple effects. Thus, squeeze activity on the incorrect side occurred more frequently when the target was flanked by letters associated with the incorrect response.

This result confirms our previous findings (O'Hara et al, 1981) and lends strong support to the continuous flow model. Evidence for the incorrect response is present in the incompatible array, and this evidence appears to be activating the incorrect response even though a correct response may be given ultimately. This activation can occur before the stimulus array is completely evaluated.

This is not the whole picture, however. Subjects also make incorrect responses and exhibit activity on the incorrect side on compatible trials, when there is nothing in the stimulus array to activate the incorrect side. This observation suggests the operation of another response-driving process that is independent of the stimulus. We label this process "aspecific priming". The analyses presented below reveal more about its nature.

Latency analysis. In the previous section, we demonstrated that there is variability both within and between conditions in the degree to which activity is present on the correct and incorrect side. We have also seen that there are differences between conditions in both traditional RT and P300 latency measures. We now evaluate the effects of the experimental manipulations on a variety of latency measures for each of the four response categories. The latency measures include: EMG and squeeze onset for the correct side, EMG and squeeze onset for the incorrect side, and P300. Figure 5 shows mean latency values for the different conditions of the experiment for each of these five latency measures. The data are segregated for the four response categories. To highlight the effects of the

frequency of trials in each category as a percentage of the total number of trials for that condition. For one subject, for some conditions, no trials were classified in the N category, and this subject's data were not considered in the subsequent analyses. For seven other subjects, the Error category was sometimes empty. The data for these subjects were retained except for one analysis (see below). An ANOVA was performed on the percent data for 11 subjects. There were four within-subject factors: the three independent variables and Category (N, E, S, or Error). Note that we have reduced the degrees of freedom associated with the Category factor from 3 to 2 because the sum of the percentages across the four categories was always 100%.

Mean percent values for the eight conditions and four categories are shown in Figure 4.

Insert Figure 4 About Here

A significant main effect of Category, $F(2, 20) = 24.89$, $p<.001$, indicated that the difference between categories was consistent: 47% of trials were classified in the N category, while only 6% of trials were classified as Errors. Corresponding values for E and S categories were 31% and 16% respectively.

Although the interaction between Warning and Category was not significant, there was a tendency for fewer trials to be classified as N, and more as S, when the warning tone was presented. We will return to this finding later. A significant interaction between Noise and Category, $F(2, 20) = 18.38$, $p<.001$, indicated that incompatible noise trials were less frequently classified as N and more frequently classified as S than

N - Activity only on the correct side in EMG and squeeze channels. (No activity on the incorrect side)

E - Activity on the correct side for EMG and squeeze channels: activity also present for EMG on the incorrect side. (EMG activity on the incorrect side)

S - Activity on the correct side for EMG and squeeze channels: activity also present for both EMG and squeeze channels on the incorrect side. The incorrect squeeze may or may not reach criterion. (Squeeze activity on the incorrect side)

Error - Activity on the incorrect side for EMG and squeeze channels. EMG activity on the correct side may or may not be present. However, no correct squeeze activity is present.

Note that there were no trials for which activity occurred in the squeeze channel, but not in the EMG channel for the same side. This is precisely what would be expected if the presence of EMG activity was intimately related to the execution of a squeeze response. In terms of a traditional error analysis, trials classified as N and E would be considered "correct" trials. On the other hand, trials classified as Error would be considered "incorrect" trials. Our S trials might be considered either "correct" or "incorrect", depending on the magnitude and timing of the two squeeze responses.

Frequency of errors. For each subject and each of the eight conditions, we determined the number of trials falling into each of the four categories described above (N, E, S, and Error) and then expressed the

apparent advantage in RT for the compatible fixed condition is not attributable to a change in stimulus evaluation time. Rather, as we have seen, RT decreases at the cost of an increase in error rate.

Error analysis

To analyze our data in more detail, we classified all the trials according to the "degree of error" apparent on each trial. Traditionally, errors are defined in a binary fashion. Subjects either do, or do not, err on a given trial. However, even when a correct response is executed on a particular trial, it is possible that an incorrect response is initiated but not completed. These partial errors will be missed if one defines the accuracy of a response in terms of the button that is pressed - or, in our case, the dynamometer that is squeezed to criterion.

The error analysis to be presented here is based on the measurement of EMG activity in the muscles associated with the incorrect response, and on the measurement of squeezes of the incorrect hand that may, or may not, reach the force criterion to be deemed "responses". It will be recalled that our subjects were required to execute a squeeze at 25% of maximum force for a response to be registered. The use of this response requirement insured that if sub-threshold response activity was present it should be observable in the form of an increase in EMG activity and/or a squeeze with less force than the 25% requirement.

We began this analysis by coding each trial in terms of the presence or absence of activity in the EMG and squeeze channels associated with correct and incorrect responses. Tabulation of trials according to these codes revealed that 99.4% of all trials could be categorized into one of the following categories:

Since we are particularly interested in stimulus evaluation processes, our ERP analysis focuses on the latency of the P300. For each trial, of each condition, P300 latencies were obtained using the technique described in the Method section. Then, an analysis of variance using the same design as that described in the previous section was used to evaluate the effects of condition on P300 latency. The results of this analysis (see Figure 3) revealed significant main effects of Noise, $F(1, 11) = 33.92$, $p<.001$, and Blocking, $F(1, 11) = 11.96$, $p<.01$.[2]

Insert Figure 3 About Here

The effects of noise (37 ms) and blocking (15 ms) were similar to the corresponding effects on RTs, whereas the effect of the warning manipulation was quite different. In fact, the presence of the warning tone did not significantly affect P300 latency, $F(1, 11) = 0.38$, $p>.05$, although it did affect RT. Given that previous studies (see above) have demonstrated that P300 latency is sensitive to those variables that affect stimulus evaluation time, we argue that the presence of a warning tone, in this experiment, speeds reaction times by affecting motor processes rather than stimulus evaluation processes. On the other hand, noise and blocking do affect stimulus evaluation time.

The latter finding supports our intepretation of the effects of blocking on RT - that is, for both compatible and incompatible arrays, stimulus evaluation is speeded under fixed compared to random conditions. Furthermore the absence of a significant Blocking by Noise interaction for P300 latency, $F(1, 11) = 2.94$, $p>.05$, suggests that the effects of noise and blocking on stimulus evaluation time are additive. Thus, the additional

error rate is larger and RT faster for the fixed than for the random condition. This suggests that subjects adopt a less conservative strategy in the blocked condition. In contrast, for incompatible noise, error rate is lower and RT faster for the fixed than for the random condition. This pattern of data cannot be readily explained in terms of a difference in the conservatism of the response criterion. Rather, it appears that the processing of the incompatible array is facilitated in fixed versus random conditions. As we will discuss later, we believe that this processing advantage is actually present for both compatible and incompatible conditions. However, it is not apparent in the compatible condition because of a concurrent change in strategy. The problem of interpretation introduced by variations in response strategy may be resolved by the P300 data to which we now turn.

ERP data

Average ERPs for the eight conditions of the RT experiment are shown in Figure 2. Note that negative going potentials are represented by an upward deflection of the curve.

Insert Figure 2 About Here

For the warned condition, we note a response to the warning stimulus followed by a slow increase in negativity (particularly at Cz) that may correspond to the contingent negative variation (CNV, Walter, Cooper, Aldridge, McCallum, & Winter, 1964). After presentation of the array, there is a "classic" P300 characterized by maximal positivity at the Pz electrode. In the unwarned condition, we also see the classic P300 following the presentation of the array.

Blocking, $F(1, 11) = 15.60$, $p<.01$. Reaction times were faster for compatible noise arrays (397 for compatible noise arrays and 444 ms for incompatible noise arrays), warned trials (410 ms for warned trials and 430 ms for unwarned trials), and fixed trial blocks (413 ms for the fixed and 428 ms for the random noise condition). The interaction between Blocking and Noise was significant, $F(1, 11) = 5.14$, $p<.05$. An analysis of simple effects revealed that the advantage for the fixed condition was more pronounced for compatible arrays (19 ms) than incompatible arrays (11 ms).

Insert Figure 1 About Here

The data relating to the effects of noise replicate those obtained by Eriksen and his colleagues. We should also note that there is an advantage of the non-informative warning stimulus of 20 ms. We shall consider the Noise x Blocking interaction when we have reviewed the error data.

Errors (defined as squeezes above the 25% force criterion with the incorrect hand) were analyzed using a similar ANOVA. For each trial block, error rate was computed as the percentage of total trials on which an error occurred. Mean error rate for the different conditions are shown in Figure 1. The error rate was larger for incompatible noise than for compatible noise trials, $F(1, 11) = 30.97$, $p<.001$. However, the effects of noise and of blocking interacted, $F(1, 11) = 34.53$, $p<.001$. In fact, separate analyses for compatible and incompatible noise revealed that fixing the level of noise for a block reduced the error rate for the incompatible noise trials, but increased the error rate for the compatible noise trials, relative to the random condition. When these data are considered together with those for RT, the following picture emerges. For compatible noise,

integrated EMG exceeded this criterion, an EMG response was deemed to have been initiated and the latency of this activity was noted. As with the squeeze responses, EMG responses in both arms could be observed on the same trial.

Results and Discussion

This section is organized in the following way. First, we present the results of an analysis of the RT and error data. This will show that we have replicated the effects reported by Eriksen and his colleagues. Then, we consider the ERP data to determine the degree to which these data are consistent with those for RT. Next, we review the results of a more fine-grained analysis in which we consider the effects of the independent variables on the frequencies of different types of errors and on the latencies of overt responses, as well as those of EMG, and ERP responses. Finally, we present speed/accuracy trade-off functions for the different conditions of the experiment as well as for different latencies of our ERP responses.

The ERP data for the count conditions will not be considered in detail. These conditions were included to confirm that any effects of the independent variables on ERP measures in the RT task could not be attributed to the motor response requirement.

Reaction time and error rate

For each subject and each of the eight conditions (defined by the three independent variables) mean RTs were derived for all correct response trials. Recall that RT was defined as the latency at which the squeeze response crossed the criterion (25% of maximum force). An analysis of variance (ANOVA) on these data (see Figure 1) revealed significant effects of Noise, $F(1, 11) = 129.59$, $p<.001$, Warning, $F(1, 11) = 44.39$, $p<.001$, and

Thus, for these trials, we were able to determine both the presence and latency of "partial" squeezes. When they occurred, these partial squeezes were generally made by the incorrect hand and were accompanied by complete overt response execution by the correct hand.

Psychophysiological data. For every trial, the variance of the EOG activity was computed. When this exceeded a preset criterion, the data from that trial were discarded. In fact, this occurred for less then 10% of the trials. The remaining single trial data from the three scalp electrodes (Fz, Cz, and Pz) were smoothed using a low pass digital filter (high frequency cut-off point at 3.14 Hz, two iterations). The three waveforms were then combined to yield a composite waveform by differentially weighting the three electrodes (Vector Filter, Gratton, Coles, & Donchin, 1983). The weights were chosen to reflect the scalp distribution usually observed for P300 (Pz>Cz>Fz). This procedure has proved to be both reliable and valid (Gratton, Kramer, & Coles, 1984; Fabiani, Gratton, Karis, & Donchin, in press). P300 latency was then estimated by finding the latency of the maximum value on the composite waveform in a time window between 300 and 1000 ms after array presentation. In this way, for each individual trial, except those where excessive eye-movements occurred, a value for P300 latency was obtained.[1]

For the respond task only, the integrated EMG activity from both arms was evaluated on each trial. To determine the latency of the onset of an EMG response, and to evaluate whether a response was present, a criterion value was established. This was accomplished using a procedure similar to that described above for the onset of squeeze activity. Thus, we determined (for each subject) the minimum value of the EMG output sufficient to discriminate a change from random variations in background EMG. When the

In each case, the derived voltage by time functions were digitized at 100 Hz, for an epoch of 2100 ms starting 1100 ms before array presentation. For the warned condition, this provided a 100 ms sample before the presentation of the warning tone.

Data Reduction

Overt responses. As we noted above, the subjects were required to squeeze the dynamometers to a criterion of at least 25% of maximum force to register a "response". Thus, an overt response was deemed to have occurred if this criterion was achieved, and RT was defined as the interval between array onset and the point at which the criterion was crossed. By evaluating the outputs of both force transducers we were able to establish both the accuracy and latency of these overt responses on every trial.

The squeeze response requirement was used to provide addional information about the dynamics of overt response execution. Thus, the output of the force transducer could be used not only to assess when the force exerted by the subject crossed the criterion, but also to determine when an overt response was initiated. In particular, we established the minimum value of output of the force transducer which was discriminable from noise. This value became the criterion for overt response initiation and the time at which this occurred was used to define the latency of squeeze onset .

In this way, for each squeeze of either dynamometer to criterion, two latency measures were available: the latency of squeeze onset and the RT. Since the outputs of both dynamometers were evaluated on each trial, these two measures were available for both correct and incorrect responses. Furthermore, on some trials, overt responses were initiated but not completed -- that is, the force exerted did not exceed the 25% criterion.

trials on which a designated central target letter was presented. For half the subjects, the counted letter was H, while for the others it was S.

On half the blocks, a warning tone (1000 Hz, 50 ms duration, 65 dB) preceded the presentation of the array by 1000 ms. These blocks constituted the warned condition. Note that the interstimulus interval (time between arrays) was the same for both warned and unwarned blocks.

For half the blocks, the level of noise (compatible or incompatible) was fixed within a block; for the other half it was random. Thus, in the fixed condition only two of the four arrays were presented, while in the random condition any one of the four arrays could occur on any trial.

Psychophysiological Recording

The electroencephalogram (EEG) was recorded from Fz, Cz, and Pz (according to the 10/20 system, Jasper, 1958) referenced to linked mastoids using Burden Ag/AgCl electrodes affixed with collodion. Vertical electrooculographic activity (EOG) was recorded from Burden electrodes placed above and below the right eye. The EMG was recorded by attaching pairs of Beckman electrodes on both the right and the left forearm using standard forearm flexor placements (Lippold, 1967). For EEG and EOG electrodes the impedance was less than 5 KOhm; for EMG, impedance was below 15 KOhm.

The EEG and EOG signals were amplified by Grass amplifiers (model 7P122), and filtered on-line using a high frequency cut-off point at 35 Hz and a time constant equal to 8 sec. for the high pass filter. The EMG signals were conditioned using a Grass Model 7P3B Preamplifier and integrator combination. The preamplifier had a 1/2 amplitude low frequency cut-off at 0.3 Hz, while the output of the integrator (full wave rectification) was passed through a filter with time constant of 0.05 sec.

with the constraint that no more than two consecutive blocks could have the same level of task, warning, or blocking.

Apparatus and Procedure

On each trial, one of four stimulus arrays, HHHHH, SSSSS, SSHSS, and HHSHH, was back-projected on a translucent screen using a Kodak random access slide projector. Stimulus duration (100 ms) was controlled by a shutter. The interval between two consecutive stimulus presentations varied randomly between 4500 and 6500 ms. The subject sat facing the screen at a distance of two meters such that the angle subtended by each letter was .5 degrees. Thus, the visual angle subtended by the entire array was 2.5 degrees. A fixation point, placed .1 degrees above the location of the central target letter, remained visible throughout the experiment.

In the respond conditions, the task of the subject was to respond to the central target letter (H or S) by squeezing one of two zero displacement dynamometers (Daytronic Linear Velocity Force Transducers, Model 152A, with Conditioner Amplifiers, Model 830A, see Kutas & Donchin, 1977). The force applied to the dynamometer was transformed into a voltage by the transducer. This voltage was digitized at 100 Hz for 1000 ms following array presentation. The output of the transducer was processed by a circuit to determine when the force exceeded a prescribed criterion value. This value defined the occurrence of an overt response and was used to determine RTs. Before the practice trials, the value of each subject's maximum squeeze force was determined for each hand separately. Then, criterion values corresponding to 25% of maximum force were established. During the practice trials, a click was presented to the subject over a loud-speaker whenever the force exerted on the transducer crossed the criterion.

In the count condition, subjects were required to count the number of

Design

Subjects were required to make a discriminative response as a function of the target letter in a five letter stimulus array. They received 12 blocks of 80 trials during each of two sessions. The first 8 blocks of the first session were considered training and the data obtained from these blocks were not used in the analysis. The remaining 1280 trials (16 blocks) were divided as follows:

Task. In half (8) of the blocks the subjects were instructed to respond with one hand to the target letter H, and with the other to the target letter S. The relationship between responding hand and target letter was counterbalanced across subjects. In the other half of the blocks the subjects were instructed to count one of the two target letters (counterbalanced over subjects).

Noise. On half the trials, the target letter was surrounded by the same letter (compatible noise), on the other half, the surrounding letters were those calling for the opposite response (incompatible noise).

Blocking. In half of the blocks, the fixed condition, only one type of noise was presented (compatible or incompatible), while in the other half, the random condition, both types of noise were presented at random. In each case, the probability of each target letter was .5.

Warning. For half the blocks, a warning tone preceded the stimulus. In the other half, no warning was given.

As a result of these manipulations, 80 trials were obtained for each of 16 conditions defined by the factorial combination of two types of task, two types of noise, two types of blocking, and two levels of warning. Note that, with the exception of noise, the level of each variable was always constant for a given block of trials. Trial blocks were randomly ordered

Donchin, 1984). Thus, the latency of P300 is sensitive to the duration of stimulus evaluation processes.

Several investigators have used measures of P300 latency as an index of stimulus evaluation time to elucidate the nature of cognitive processes (e.g. Brookhuis, Mulder, Mulder, & Gloerich, 1983; Duncan-Johnson & Kopell, 1981; Ford, Roth, Mohs, Hopkins, & Kopell, 1979). In the present study, we used P300 latency to determine whether the noise/compatibility manipulation influences stimulus evaluation time. As in the O'Hara et al. (1981) study, we measured EMG from the limbs associated with both correct and incorrect responses. We also required responses (squeezes of zero-displacement dynamometers) that provided an additional level of measurement of response activation, namely, squeezes that did not reach criterion. Using these psychophysiological measures we could evaluate the effects of the noise/compatibility manipulation on both stimulus evaluation and response competition. Finally, we were interested in the role of preparatory processes in the Eriksen paradigm. On some trials, a warning tone preceded the presentation of compatible and incompatible arrays by 1000 ms. If, as Posner (1978) has argued, the effect of this type of alerting stimulus is to influence motor preparation, then P300 latency to the arrays should not be influenced by the warning tone. Rather, RT may be shortened at the expense of error rate.

Method

Subjects

Twelve male students at the University of Illinois (aged between 18 and 23) served as subjects. They were paid $3.50 per hour, plus a bonus for participating in all sessions.

noise/compatibility and warning manipulations, we present the latency data for these manipulations in Figures 6 and 7, respectively. The latter two figures also provide information about the frequency of the different response categories for the two manipulations.

Insert Figure 5 About Here

Insert Figure 6 About Here

Insert Figure 7 About Here

---

a. Latency of correct activity. For the N, E, and S categories, correct activity is present in both EMG and squeeze channels. Two separate ANOVAs were used to determine whether the latency of activity in these two channels was affected by the experimental conditions and varied as a function of response category. Significant main effects of Warning, Blocking, Noise, and Category were evident for the latencies of both EMG and squeeze activity (see Table 1).

Insert Table 1 about here

The direction of these effects was the same as found in the RT analysis described above. Activity in these two channels occurred earlier on

compatible versus incompatible trials, on warned versus unwarned trials, and for fixed versus random trial sequences. Furthermore, individual comparisons between pairs of adjacent categories indicated that the latency of both aspects of correct activity increased as a function of the degree of incorrect activity present (i.e. from N to E to S categories). These effects can be seen in Figures 5, 6, and 7.

Further scrutiny of these figures suggests that the time between the onset of correct EMG activity and the onset of correct squeeze activity varies with error - that is, as the amount of incorrect activity increases (from N to E to S), there appears to be some interruption in the execution of the correct response. This suggestion was confirmed by an ANOVA on the values for the difference in latency between correct squeeze and correct EMG onsets, $F(2, 20) = 32.30$, $p<.001$. Comparisons between individual response categories revealed that this difference was largest for the S category, while N and E categories did not differ significantly from each other.

Thus, when incorrect activity is present in any form, the onset of correct activity is delayed. In addition, <u>when incorrect squeeze activity is evident, there is a further delay in the execution of the correct response</u> - that is, the time between correct response initiation (as indicated by EMG activity) and completion (as indicated by the squeeze) is prolonged. This delay in correct response execution is evidence for a response competition mechanism which is responsible, at least in part, for the difference in overall RT between compatible and incompatible noise arrays. Recall that fewer trials are classified in the N category, and more trials in the S category, for incompatible than for compatible arrays (see Figure 6). Since correct response latencies are shorter for N than for S categories, measures of mean RT (without regard to category) will

necessarily be longer for incompatible arrays.

This response competition process is apparently not the only factor responsible for the noise/compatibility effect. When response latencies are evaluated for the same level of response competition, (i.e. when category is taken into account) a compatibility effect is still visible (see Figure 6). We will return to this point following a discussion of the P300 latency data below.

The analysis of the differences between the latencies of correct EMG and correct squeeze activity also revealed a significant difference between compatible and incompatible noise arrays, $F(1, 10) = 5.31$, $p<.05$ (see Figure 6). The relevant means were 59 ms for compatible and 67 ms for incompatible arrays. Following the response competition arguments above, we interpret this difference as indicating greater response competition when the target letter is flanked by incompatible letters. However, this difference is independent of category - that is, it is constant over categories and present even in the N category, where, by definition, we failed to pick up any external manifestation of incorrect activity. It is possible that response competition effects may occur either at a level of response activation that precedes EMG activity or at a level of EMG activity that is not detected by our procedures.

b. Latency of incorrect activity. For E, S, and Error categories, EMG activity is evident on the incorrect side, while for S and Error categories, incorrect squeeze activity is also apparent. The next series of analyses evaluates the latency of this incorrect activity as a function of the experimental conditions and of response category. This analysis is complicated by the fact that not all subjects have trials in all categories for all conditions. Eleven subjects have sufficient trials in the E and S

categories for all conditions, but only four subjects have trials in the Error category for all conditions. Therefore, we performed two sets of overlapping analyses. The first, based on data for eleven subjects, covers the E and S categories: the second, based on the data from four subjects, relates to the E, S, and Error categories.

Both analyses indicated that the latency of incorrect EMG activity varies significantly with category (see Table 1, and Figure 5). For the analysis based on 11 subjects, EMG latency is shorter for S (355 ms) than E (388 ms) categories; for the analysis based on four subjects, EMG latency was shortest for the Error category (300 ms), intermediate for the S category (349 ms), and longest for the E category (404 ms). Tukey's HSD test (Tukey, 1953) confirmed that the latencies of the Error and E categories were significantly different, while neither category was statistically distinguishable from the S category. For the analysis based on four subjects, incorrect squeeze latency was shorter for the Error category (368 ms) than the S category (418 ms). The mean incorrect squeeze latency for the S category for the larger group of 11 subjects was 396 ms.

It is apparent that the <u>latency</u> of incorrect activity decreases as the <u>degree</u> of incorrect activity increases (from E to S to Error). Note that the subset of 4 subjects had mean incorrect response latencies that are quite similar to those for the larger group of 11 subjects.

The analysis based on 11 subjects revealed significant condition effects on the latencies of incorrect activity. For incorrect EMG and squeeze channels, latencies were significantly shorter on warned trials and for compatible noise arrays (see Table 1 and Figure 5). These effects were not significant when the same analysis was performed on the data from the subset of 4 subjects, presumably because of insufficient power. In each

case, however, the means were ordered in the same direction as those for the larger analysis.

When the results of the analyses of both incorrect and correct activity are considered together, several conclusions may be drawn. First, as noted earlier, there appears to be competition between correct and incorrect responses such that correct responses are delayed if incorrect activity is present. We argued above that this competition effect is partly responsible for the differences in mean RT between compatible and incompatible arrays (see Figures 5 and 6). Second, we have seen that correct activity is delayed and incorrect activity is speeded as the degree of error increases. Furthermore, the latencies of both correct and incorrect activity (EMG and squeeze) are shorter when a warning tone precedes the presentation of the stimulus array (see Figures 5 and 7). These data suggest the presence of an additional process that influences both the latency of response and its correctness. This is the process of "aspecific priming" that we mentioned earlier.

As its name implies, "aspecific priming" refers to a response activation process that occurs without regard to the nature of the stimulus. We use the word "priming" to indicate activation of response channels and "aspecific" to indicate that this activation is not controlled by the specific information provided by the current stimulus. We propose that, on a trial to trial basis, either or both responses (with the left and/or right hand) can be primed in advance of stimulus presentation, or at least activated independently of the nature of the stimulus presented. The degree to which the subject makes an error (shows incorrect activity) depends in part on the level of priming of the incorrect response. When the level is high, minimal information can trigger the incorrect response. This

information may be provided by the mere presentation of a stimulus (as proposed by Grice, Nullmeyer, & Spiker, 1982) and/or the initial evaluation of the features in the stimulus array. The first mechanism is responsible for the emission of incorrect responses for compatible arrays, as well as for the larger number of errors for the fixed compatible and warned conditions. It is the mechanism that leads to a "fast guess". The second mechanism is responsible for the larger frequency of errors for the incompatible arrays. The probability that aspecific priming of the incorrect response will be manifested in an overt response will depend on the level of the activation. This, in turn, will determine whether, once the response has been initiated, it can be overriden and the correct response produced. If the incorrect response is initiated soon after stimulus presentation, it cannot be countermanded and an Error trial occurs. If the incorrect response is initiated a little later (there is less activation of the incorrect response) an S trial will occur. And so on.

ERP data. Measures of P300 latency were available for each trial. Thus, it was possible to conduct a series of analyses to evaluate the relationships between this latency measure and both the experimental manipulations and response category. These analyses paralleled those described in previous sections for measures of motor response latency. Again, two overlapping analyses were performed. The first involved the data from 11 subjects and covered the N, E, and S categories; the second focused on the data for the subset of four subjects and covered all four categories (i.e. N, E, S, and Error).

The first analysis revealed significant main effects of Blocking, Noise/Compatibility, and Category (see Table 1). P300 latencies were shorter for the fixed (602 ms) than for the random (616 ms) condition, and

for compatible (593 ms) than for incompatible (625 ms) noise arrays Furthermore, longer P300 latencies were observed for those categories in which there was activity in the incorrect channels. The relevant means were as follows: for N, 591 ms; for E, 601 ms; and for S, 635 ms. Separate comparisons among the three response categories revealed that the effect was due to an increase in P300 latency for the S category. The N and E categories were not statistically distinguishable. Figure 5 provides a representation of these effects broken down by conditions. Figures 6 and 7 show the mean effects of the noise and warning manipulations.

The second analysis, on the data for the subset of four subjects, revealed a significant effect of Category (see Table 1). P300 latency increased across category. The relevant means were as follows: for N, 623 ms; for E, 631 ms; for S, 646 ms; and for Error, 707 ms. Tukey HSD tests indicated that the Error category was associated with a later P300 than N and E categories. No other significant main effects were obtained. As above, we attribute the difference between the two analyses to lack of power, since the means for the subset analysis were ordered in the same way as those for the analysis based on 11 subjects.

Taken together, these data suggest that stimulus evaluation processes are longer when subjects are presented with incompatible noise arrays and when the the level of noise is randomized. In addition, the duration of the stimulus evaluation process is related to the likelihood of incorrect activity, as manifested either in the EMG or squeeze channels.

It is interesting that P300 latency was not affected by the presentation of a warning stimulus (see Figure 7). This suggests that the effects of warning on response latency measures are not due to stimulus evaluation differences. Rather, it appears that subjects initiate responses

earlier within the stimulus evaluation process when they are forewarned. However, the duration of the process preceding the emission of the P300 is not changed by the warning or by the early emission of the response. In fact, an analysis of the difference between P300 and correct EMG latencies revealed that, on warned trials, EMG activity began 235 ms before the P300 occurred, while on unwarned trials, this difference was 211 ms, $F(1, 10) = 10.22$, $p<.01$. We propose that this effect is due to greater aspecific priming - that is, in warned conditions, subjects activate their response systems to a greater extent in advance of stimulus presentation because they can time this activation to coincide with stimulus presentation.[3] This results in faster RTs but more trials with incorrect activity.[4]

Speed/Accuracy Trade-Off Functions

An alternative method of evaluating these data is to consider speed-accuracy trade-off functions. These functions are obtained by plotting response accuracy as a function of response latency. They are intended to provide a representation of the manner in which stimulus evaluation processes proceed over time that is uncontaminated by response bias factors (e.g., Pachella, 1974). However, this interpretation is predicated on the assumption that the speed of stimulus evaluation processes is constant for a given condition. This assumption may not be valid (see Meyer & Irwin, 1982). The speed/accuracy functions we present here have P300 latency as a parameter. That is, trials are first sorted according to the latency of the P300. Then, for each P300 latency bin, we plot response accuracy against RT. If P300 latency can be taken as a measure of the duration of stimulus evaluation, then we have a series of speed-accuracy trade-off functions with stimulus evaluation duration as an independent parameter.

The functions we present below provide the following information. First, we show that speed-accuracy trade-off functions are indeed different for trials aggregated according to P300 latency. Second, we show how two of our manipulations, noise/compatibility and warning, have a different effect on these functions. Finally, we use functions derived for each experimental condition and different P300 latencies to gain insight into the process of stimulus evaluation.

We obtained our functions in the following way. For each of the 12 subjects, and for each of the eight conditions, the latency of the first squeeze response (RT), the correctness of that response, and the P300 latency for each trial were tabulated. Then, for RT bins of <350 ms, 350-449 ms, and 450-549 ms, accuracy estimates were computed by dividing the number of correct trials in a bin by the total number of trials in the same bin. This procedure was performed separately for trials on which P300 latency was longer, and shorter, than the median P300 latency for that subject and condition. The values of the RT bins were chosen to encompass the range of RTs exhibited by the subjects. Some of the 576 cells (12 subjects x 8 conditions x 2 P300 latencies x 3 RT bins) did not have a sufficient number of trials to obtain accuracy estimates (less than 3 trials). Thus, we did not perform analyses of variance on the accuracy data. Instead, we computed the means (over subjects) and associated standard errors.

Figure 8 displays a summary of the speed-accuracy trade-off functions for different P300 latency, noise, and warning conditions. Figure 9 gives the 16 functions on which these summaries were based. The standard errors for each mean are also shown in the figure.

Insert Figure 8 About Here

In Figure 8a, we note that, regardless of the latency of the P300 (i.e. the duration of stimulus evaluation), accuracy increases as RT increases -- that is, the slower the response the more likely is the subject to be correct. However, accuracy is lower for all response speeds when P300 latency is long. Furthermore, the same level of accuracy is achieved either by the conjunction of a slow RT and a slow P300, or by a fast RT and a fast P300 - that is, P300 latency, and by implication stimulus evaluation time, determines the relative position of the speed-accuracy trade-off function. These data indicate that the accuracy of a response depends on its timing relative to the evaluation process. When evaluation proceeds quickly, a high level of accuracy is achieved even when RTs are short: conversely, when evaluation proceeds slowly, a high level of accuracy is only achieved when RTs are long. These data illustrate how measures of the P300 can be used to overcome the difficulties raised by the assumption that the duration of the evaluation process is constant on every trial.

Figure 8b shows the speed-accuracy functions for compatible and incompatible noise arrays. Note that, for each RT bin, accuracy is lower for the incompatible arrays. This confirms that the evaluation process is slower, or at least different, for these arrays.

Figure 8c shows the functions for warned and unwarned trials. These functions are essentially identical. This observation confirms the conclusion we drew earlier that the presence of a warning stimulus does not affect the evaluation process. Rather the difference between these two

conditions in mean response latencies and error rates reflects a difference in the average point on the speed-accuracy tradeoff function at which the subject is operating. As we argued above, the greater aspecific priming on warned trials leads to a less conservative response (i.e., responses are released on the basis of less information).

Figure 9 shows the speed/accuracy trade-off functions for different P300 latencies for the 8 conditions separately.

Insert Figure 9 About Here

The most interesting aspect of these functions concerns the accuracy for fast reaction times and slow P300s. In the compatible noise conditions, accuracy is approximately 50%. We infer from this that when the subject responds quickly on trials where the duration of stimulus evaluation is long (P300 latency is long), he is essentially guessing. However, on incompatible trials, the combination of fast RTs and slow P300s is associated with an accuracy value that is below chance.

One explanation for this excessive error rate is that, early in the evaluation of an incompatible noise array, there is more evidence for the incorrect response. It should be recalled that an incompatible array contains one letter associated with the correct response and four letters associated with the incorrect response. Thus, when the subject responds quickly and evaluation is proceeding slowly, the evidence available at the time of response favors the incorrect response. Note that this excessive error rate is not seen in the data for compatible arrays. Our data suggest, then, that early in the evaluation process, the subject performs an analysis of the features of all the letters in the array, without selecting the

information provided by the target letter in the central location. We refer to this process as "feature", or "letter", analysis. Selection for the features of the center letter ("location" analysis) appears to occur later. These two aspects of stimulus evaluation, feature, or letter, analysis and location analysis, can both activate the response channels directly. The two processes may occur in sequence or in parallel. However, in the latter case, feature analysis should be faster than location analysis. Thus, early responses, based mainly on the feature analysis, are likely to be incorrect for an incompatible noise trial, but correct for a compatible noise trial.

The process of aspecific priming, discussed earlier, also controls activation of response channels. If one or other of the responses is heavily primed (for example, because of guessing), then that response may be released without being influenced by either feature or location analyses.

Conclusions

The results of this experiment are clearly consistent with the continuous flow model of information processing (Eriksen & Schultz, 1979). We have found that the correct and the incorrect response channels can be activated concurrently. This activation occurs either as a result of the evaluation process <u>and/or</u> because of aspecific priming. In the former case, as evaluation proceeds and before it is completed, information is accumulated about all the letters in the array, and this information is fed continuously to the activation system. When the array contains incompatible noise, this information will call for the incorrect response. In the case of aspecific priming, either or both responses can be primed prior to stimulus presentation. Alternatively, the presentation of the array leads to an activation of responses independent of the nature of the stimulus. Note that we propose the existence of a <u>specific</u> activation process, driven

by stimulus evaluation, to account for the fact that incorrect responses are activated to a greater extent for incompatible noise than for compatible noise arrays. Aspecific priming is necessary to account for the presence of incorrect responses to compatible arrays when there is nothing in the stimulus itself to drive an incorrect response.

While our data are consistent with some kind of continuous flow model, they are not easily accommodated by a strictly serial stage model. It is difficult to see how a serial model can account for the concurrent activation of both correct and incorrect responses. The model cannot readily encompass the observation that early responses to incompatible arrays are generally incorrect.

The analysis of the EMG and sub-threshold squeeze data have important implications for the concept of response competition. First, we find that when incorrect activity is present, initiation of correct activity is delayed progressively. Since the P300 is delayed in a similar fashion, it appears that the probability of response competition increases as the time for stimulus evaluation increases (cf. Miller, 1983). Second, we find that the temporal characteristics of correct response execution are affected by the degree to which incorrect activity is present. When an incorrect squeeze response is produced (the S category), the time between correct EMG onset and correct squeeze onset is increased. This finding is most readily explained in terms of the operation of a response competition mechanism. The fact that the temporal characteristics of response execution can be modified, and responses can be initiated without being executed, suggest that response execution is best conceived of as a continuous process. This view contrasts with that of McClelland (1979), for whom response execution is the only discrete process in the human information processing system.

The manipulations we used in our experiment have different effects on the information processing system. One effect of introducing incompatible noise to the stimulus array is to increase the number of trials on which incorrect activity occurs. In general, the presence of incorrect activity is associated with an increase in the time taken to execute a correct response. Thus, the mean RT difference between compatible and incompatible noise is due, at least in part, to response competition. However, the effect of incompatible noise is also to slow down the evaluation process, as indexed by P300 latency. Furthermore, correct EMG and squeeze latencies are longer for incompatible than for compatible noise, even when response competition effects are controlled. Thus, the noise/compatibility effect on mean RT appears to be due both to an effect on the incidence of response competition and to an effect on the stimulus evaluation process.

In contrast to the noise manipulation, the warning conditions provided a clear dissociation between P300 latency and the latency of response measures (correct and incorrect squeeze and EMG onset latencies). The latter were in fact shortened by the warning, while the presence of a warning had no effect on P300 latency. This result suggests that the warning did not influence stimulus evaluation processes, while it was clearly effective in increasing the aspecific priming of the two response channels. These data contrast in an interesting manner with the results of Duncan-Johnson and Donchin (1982). These investigators presented imperative stimuli that either matched, or failed to match, an antecedent warning stimulus. When the stimuli mismatched, the P300 latency to the imperative stimulus increased. Thus, there are conditions in which the information carried by a warning stimulus can affect the duration of stimulus evaluation processes for a subsequent event, suggesting the operation of perceptual

Table 1

Results of Analyses of Variance on Latency Measures

Independent variables		Dependent variables					
Main effects	Response side	EMG latency		Squeeze latency		P300 latency	
		df	F	df	F	df	F
Noise	Incorrect	1,10	8.87*	1,10	17.94*	1,10	26.44**
	Correct	1,10	17.13**	1,10	67.67**		
Warning	Incorrect	1,10	5.78*	1,10	9.32**	1,10	1.00
	Correct	1,10	8.81*	1,10	16.44**		
Blocking	Incorrect	1,10	4.75	1,10	2.90	1,10	12.19**
	Correct	1,10	6.32*	1,10	4.98*		
Response Category	Incorrect	1,10	26.60**				
		2,6	9.10* (a)	1,3	23.47* (a)		
						2,20	17.13**
						3,9	10.61** (a)
	Correct	2,20	51.80**	2,20	109.24**		

** $p < .01$ (a) analysis based on 4 subjects

* $p < .05$

of the data is the lack of a significant effect of warning on error rate. Scrutiny of Figure 8c indicates that a 20 ms decrease in reaction time (the mean effect of warning) should be associated with an increase in error rate of approximately 3%. This was, in fact, the increase in error rate when computed using the definition of an error described in this section. Because error rate was computed on a relatively small number of trials, our estimate was not sufficiently reliable to permit a 3% difference to be significant in an ANOVA. If more reliable estimates were obtained, we could determine whether the difference is "real" or whether, in fact, the subjects are able to respond faster, but at the same accuracy level, when a warning is present. If this is the case, then the effect of the warning might be to change the slope of the response activation function, that is, to speed motor processes.

Footnote

[1] We should note that we also used a more traditional method, peak-picking at Pz, to determine the latency of P300 on single trials. There was a close correspondence between the data obtained using the traditional procedure and those from vector filter. However, analyses of variance on latency measures derived from the vector procedure yielded consistently higher F values than those based on the peak-picking procedure.

[2] A similar analysis of P300 latency for the count task, when no motor response was required, also revealed a significant main effect of noise, $F(1, 11) = 11.90$, $p<.01$. As in the RT task, P300 latency was longer for incompatible arrays.

[3] The aspecific priming process has no obvious psychophysiological manifestation. However, several investigators have described lateralized scalp negativities related to motor preparation (see Deecke, Bashore, Brunia, Grunewald-Zuberbier, Grunewald, & Kristeva, 1984, for a review). Furthermore, some researchers (e.g. Rohrbaugh, Syndulko, & Lindsley, 1976; Gaillard, 1977; Kok, 1978; Rohrbaugh & Gaillard, 1983) have argued that later aspects of the CNV, which is apparent in our warned condition, are related to motor preparation. However, whether the negativity observed in our study is a manifestation of aspecific priming remains to be determined. In particular, we need to evaluate scalp activity at lateral recording sites rather than at the midline. For a review of the relation between event-preceding negativities and response preparation, see Donchin, Coles, & Gratton (1984).

[4] We have argued that the presence of a warning tone does not affect evaluation process. Rather it leads subjects to become less conservative - they respond faster and make more errors. One apparently troubling aspect

Authors' Note

Some of the data reported in this study were presented at the Joint EEG Society/Psychophysiological Society Meeting, Bristol (England), 1983, at the 7th Evoked Potentials International Conference, Florence (Italy), 1983, and in Coles, M. G. H., & Gratton, G. (in press) "Psychophysiology and contemporary models of human information processing", to appear in D. Papakostopoulos et al. (Eds.), "Clinical and Experimental Neuropsychophysiology", Beckenham, England: Croom Helm Ltd.

This study was supported in part by a contract from Air Force Office of Scientific Research, contract #F49620-83-0144, Dr. Al Fregly, Project Director. We wish to thank Rich Carlson, Demetrios Karis, Art Kramer, Mick Rugg, and Erik Sirevaag for their helpful comments on preliminary versions of the manuscript.

Ted Bashore is now at the Medical College of Pennsylvania at EPPI, Philadelphia, Pa.

Send reprint requests to:

Dr. Michael G. H. Coles
University of Illinois
Psychology Department
603 East Daniel
Champaign, Illinois 61820

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), Attention and Performance II (pp. 276-315). Amsterdam: North-Holland.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. Cognitive Psychology, 12, 97-136.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), Attention and Performance VI (pp. 333-361). Hillsdale, NJ: Erlbaum.

Tukey, J. W. (1953). The problem of multiple comparisons. Princeton University: Ditto.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. Psychological Review, 80, 1-52.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. Nature, 203, 380-384.

Yeh, Y-Y., & Eriksen, C. W. (1984). Name codes and features in the discrimination of letter forms. Perception and Psychophysics, in press.

Meyer, D. E., Yantis, S., Osman, A., & Smith, J. E. K. (1984). Discrete versus continuous models of response preparation: A reaction time analysis. In S. Kornblum & J. Requin (Eds.), Preparatory states and processes (pp. 69-94). Hillsdale, NJ: Erlbaum.

Miller, J. (1982). Discrete versus continuous stage models of human information processing: In search of partial output. Journal of Experimental Psychology: Human Perception and Performance, 8, 273-296.

Miller, J. (1983). Can response preparation begin before stimulus recognition finishes? Journal of Experimental Psychology: Human Perception and Performance, 9, 161-192.

O'Hara, W. P., Morris, L. R., Coles, M. G. H., Eriksen, C. W., & Morris, N. M. (1981). Stimulus incompatibility and response competition: An EMG/RT analysis. Psychophysiology, 18, 170-171 (Abstract).

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B.H. Kantowitz (Ed.), Human Information Processing: Tutorials in Performance and Cognition (pp.41-82). Hillsdale, NJ: Erlbaum.

Posner, M. I. (1978). Chronometric explorations of mind. Hillsdale, NJ: Erlbaum.

Rohrbaugh, J. W., & Gaillard, A. W. K. (1983). Sensory and motor aspects of the contingent negative variation. In A. W. K. Gaillard and W. Ritter (Eds.), Tutorials in Event-Related Potential research: Endogenous components (pp. 269-310). Amsterdam: North-Holland.

Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1976). Brain components of the contingent negative variation in humans. Science, 191, 1055-1057.

Kutas, M., & Donchin, E. (1977). The effect of handedness, of responding hand, and of response force on the contralateral dominance of the readiness potential. In J. Desmedt (Ed.), Attention, voluntary contraction and event-related cerebral potentials (pp. 189-210). Basel: Karger.

Kutas, M., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potentials. Brain Research, 202, 95-115.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792-795.

Lippold, O. C. J. (1967). Electromyography. In P. H. Venables and I. Martin (Eds.), A manual of psychophysiological methods (pp. 245-297). Amsterdam: North Holland.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21, 171-186.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. Science, 211, 77-80.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review, 86, 287-330.

Meyer, D. E., & Irwin, D. E. (1982). On the time course of rapid information processing . (Tech. Rep. No. 43). Ann Arbor: University of Michigan, Cognitive Science Program.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). Filtering for scalp distribution: A new approach (Vector filter). Psychophysiology, 20, 443-444 (Abstract).

Gratton, G., Kramer, A. F., & Coles, M. G. H. (1984). A comparative study of measures of the latency of event-related brain components. Psychophysiology, 21, 578-579 (Abstract).

Grice, G. R., Canham, L., & Shafer, C. (1982). Development of associative and perceptual interference. Perception and Psychophysics, 32, 375-387.

Grice, G. R., Nullmeyer, R., & Spiker, V. A. (1982). Human reaction times: Toward a general theory. Journal of Experimental Psychology: General, 111, 135-153.

Grossberg, S. (1982). Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control. Boston: Reidel Press.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography and clinical Neurophysiology, 10, 371-375.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and memory: Individual differences in the von Restorff effect. Cognitive Psychology, 16, 177-216.

Kok, A. (1978). The effect of warning stimulus novelty on the P300 and components of the contingent negative variation. Biological Psychology, 6, 219-233.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of the interference. Science, 214, 938-940.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. Perception and Psychophysics, 16, 143-149.

Eriksen, C. W., & Eriksen, B. A. (1979). Target redundancy in visual search: Do repetitions of the target within the display impair processing? Perception and Psychophysics, 26, 195-205.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. Perception and Psychophysics, 25, 249-263.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (in press). P300: Theoretical and methodological issues. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), Advances in Psychophysiology: Vol. 2. Guilford, CT: JAI press.

Flowers, J. H., & Wilcox, N. (1982). The effect of flanking context on visual classification: The joint contribution of interactions at different processing levels. Perception and Psychophysics, 32, 581-592.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, W. F., & Kopell, B. S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and clinical Neurophysiology, 47, 450-459.

Gaillard, A. W. K. (1977). The late CNV wave: Preparation versus expectancy. Psychophysiology, 14, 563-568.

References

Brookhuis, M. A., Mulder, G., Mulder, L. J. M., & Gloerich, A. B. M. (1983). The P3 complex as an index of information processing: The effects of response probability. Biological Psychology, 17, 277-296.

Chase, W. G. (1984). The timing of mental acts. In E. Donchin (Ed.), Cognitive psychophysiology: Event-related potentials and the study of cognition (pp. 221-247). Hillsdale, NJ: Erlbaum.

Deecke, L., Bashore, T., Brunia, C. H. M., Grunewald-Zuberbier, E., Grunewald, G., & Kristeva, R. (1984). Movement-associated potentials and motor control. In R. Karrer, J. Cohen, & P. Tueting (Eds.), Brain and Information: Event-Related Potentials (pp. 398-428). New York: The New York Academy of Sciences.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked potentials and behavior (pp. 13-75). New York: Plenum Press.

Donchin, E. (1981). Surprise! ... Surprise? Psychophysiology, 18, 493-513.

Donchin, E., Coles, M. G. H., & Gratton, G. (1984). Cognitive psychophysiology and preparatory processes: A case study. In S. N. Kornblum & J. Requin (Eds.), Preparatory states and processes (pp. 155-178). Hillsdale, NJ: Erlbaum.

Donders, F. C. (1969). On the speed of mental processes. In W. G. Koster (Ed. and trans.), Attention and Performance II (pp.412-431). Amsterdam: North-Holland.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. Biological Psychology, 14, 1-52.

responses without affecting stimulus evaluation time (P300 latency). It was also the case that a large negative slow wave was present in the recordings of scalp-electrical activity in the warned conditions; this wave was absent or smaller in the unwarned conditions. As we have noted, negative going potentials of this kind can precede motor response execution, and their amplitudes are related to response output requirements (Kutas & Donchin, 1977, 1980). For this reason, we prefer to conceive of the warning stimulus as leading to a greater degree of advanced motor preparation, that is activation of response channels, rather than to a change in response criterion.

In summary, the results of our experiment support the predictions of parallel models and are not congruent with strictly serial models. The data are also quite consistent with the continuous flow model (Eriksen & Schultz, 1979), although they are not inconsistent with other parallel models, such as those proposed by Miller (1982) or Grice and his colleagues (Grice, Nullmeyer, & Spiker, 1982). We have provided evidence for two relatively independent sources of response activation: an "aspecific", stimulus-independent process, and a "specific", stimulus-dependent process. As evidence accumulates in the stimulus evaluation system, specific activation of the associated response systems occurs. Activation of the incorrect channel is determined both by the amount of aspecific priming and by the evaluation process, when there is evidence in the stimulus for the incorrect response. Activation of the incorrect response channel can interfere with correct response execution through a response competition process.

stimulus array. Furthermore, at the level of feature (or letter) analysis, several grains must be handled in parallel. On the other hand, at the level of localization analysis information may be transferred in only one grain.

Our data also suggest a complex picture of response activation processes, including a "specific", stimulus driven activation process, an "aspecific", stimulus independent, priming process, and a response competition mechanism. Furthermore, we have argued that the activation of responses occurs in a continuous rather than discrete fashion. We have already discussed the specific activation process and response competition, since they are both included in Eriksen's model. On the other hand, the aspecific priming process is not explicitly included in this model, although Eriksen and Schultz (1979) do argue that instructions, set, expectancy, and pay-off schedules may pre-prime one or other of the responses so that they have a lower threshold of evocation. Furthermore, they propose that the mere presentation of the stimulus will prime a wide range of responses. Similar arguments are presented by Grice and his colleagues (Grice, Nullmeyer, & Spiker, 1982). However, for the latter authors, whether and when response channel activation will lead to an overt response depends on a response criterion (bias) which can vary on a trial-to-trial basis. We prefer to conceive of a set of relatively fixed response criteria for EMG onset, squeeze onset, etc. Variability in response latency is determined in our case by variability in initial levels of activation, produced either by priming or by activation associated with stimulus onset.

Although the two proposals, variability in criterion or variability in initial activation, produce the same predictions from an operational point of view, psychophysiological evidence suggests that our approach may be more appropriate. Recall that the effect of the warning signal was to speed up

arrays because the localization process can be by-passed. In the case of incompatible arrays, localization of the target letter may be facilitated if it is consistently presented in the context of different letters.

Both feature (letter) and localization processes appear to activate the response channels directly. In fact, the speed-accuracy functions for incompatible arrays reveal that early responses are driven more by the lateral letters than by the central target letter. This short-cut of the information processing flow is inconsistent with the assumptions of a strictly serial <u>and</u> a strictly cascade model (e.g. McClelland, 1979). Both these models assume that the flow of information proceeds through an ordered sequence of processing elements. On the other hand, these kinds of short cuts are not inconsistent with the assumptions of the continuous flow model (Eriksen & Schultz, 1979).

An interesting integration of serial and parallel models has been proposed recently by Miller (1982, 1983). His model can be described as a hybrid "parallel-discrete" model. He suggests that information is <u>not</u> transferred continuously between processing elements. Rather, the transfer only occurs when an element has completely processed a "grain" of information. Thus, information represented by a grain is transferred discretely. However, when there is more than one grain, different processing elements can be engaged in parallel. Note that, when all the relevant information is contained in one grain, his model is formally equivalent to a serial model. When the relevant information can be partitioned into an infinite number of grains, his model is formally equivalent to a cascade model. In terms of Miller's model, our data suggest that the information is partitioned into more than one grain, because responses are activated on the basis of partial information about the

priming. However, in the present study, the warning stimulus (a tone) did not match the imperative stimuli (letters). Under these circumstances, there is apparently no opportunity for an effect of perceptual priming on the evaluation process.

By blocking the level of noise, only the correct responses were speeded, indicating a facilitation of the stimulus evaluation processes. Converging evidence for this facilitation was provided by P300 latency data. A shorter latency was observed for fixed than for random conditions. Fixing the level of noise may also lead to a modification in the response criterion for compatible arrays, such that subjects respond faster but less accurately.

We conceive of the stimulus evaluation process in our experiment as consisting of at least two sub-processes, feature or letter analysis and localization analysis. Note that our conception of the process of stimulus evaluation is similar to that discussed by Treisman and her colleagues (Treisman & Gelade, 1980; Treisman, Sykes, & Gelade, 1977). They argue that an early, parallel process of feature analysis precedes the detection of the feature location. Our data suggest that these two sub-processes may occur in sequence or in parallel, although the output of the feature analysis should be available before that of the location analysis.

This analysis of the components of the evaluation process helps clarify the effects of noise/compatibility and blocking manipulations on the information processing system. In particular, the delay in the evaluation process for incompatible arrays may be explained by the conflict between the outputs of the feature (letter) and localization analyses. This conflict is not present for compatible noise arrays. Fixing the level of noise for a block of trials may facilitate the processing of compatible

Figure Captions

Figure 1. Reaction times and error rates as a function of noise, warning, and blocking conditions.

Figure 2. ERP waveforms (averaged over subjects) for three electrode locations, Fz, Cz, and Pz. Separate waveforms are shown for the eight different experimental conditions.

Figure 3. Mean P300 latencies as a function of the eight experimental conditions for the respond task.

Figure 4. Frequency distributions of trials as a function of the four response categories. Separate distributions are shown for the eight different experimental conditions.

Figure 5. Values for the five latency measures as a function of response category and the eight conditions of the experiment.

Key.

P300: Latency of the P300

Csq: Latency of onset of the correct squeeze response

Cemg: Latency of onset of the correct EMG response

Isq: Latency of onset of the incorrect squeeze response

Iemg: Latency of onset of the incorrect EMG response

Figure 6. Latency of correct EMG and squeeze activity and of P300 as a function of the degree of incorrect activity for compatible and incompatible arrays. The relative frequencies of each response category for compatible and incompatible arrays are shown in the upper panel.

Figure 7. Latency of correct EMG and squeeze activity and of P300 as a function of the degree of incorrect activity for warned and not warned trials. The relative frequencies of each response category for warned and not warned trials are shown in the upper panel.

Figure 8. Speed/accuracy trade-off curves as a function of P300 latency (8a), noise (8b), and warning (8c).

Figure 9. Speed/accuracy trade-off curves as a function of P300 latency for compatible and incompatible noise trials, for warning and blocking conditions separately.

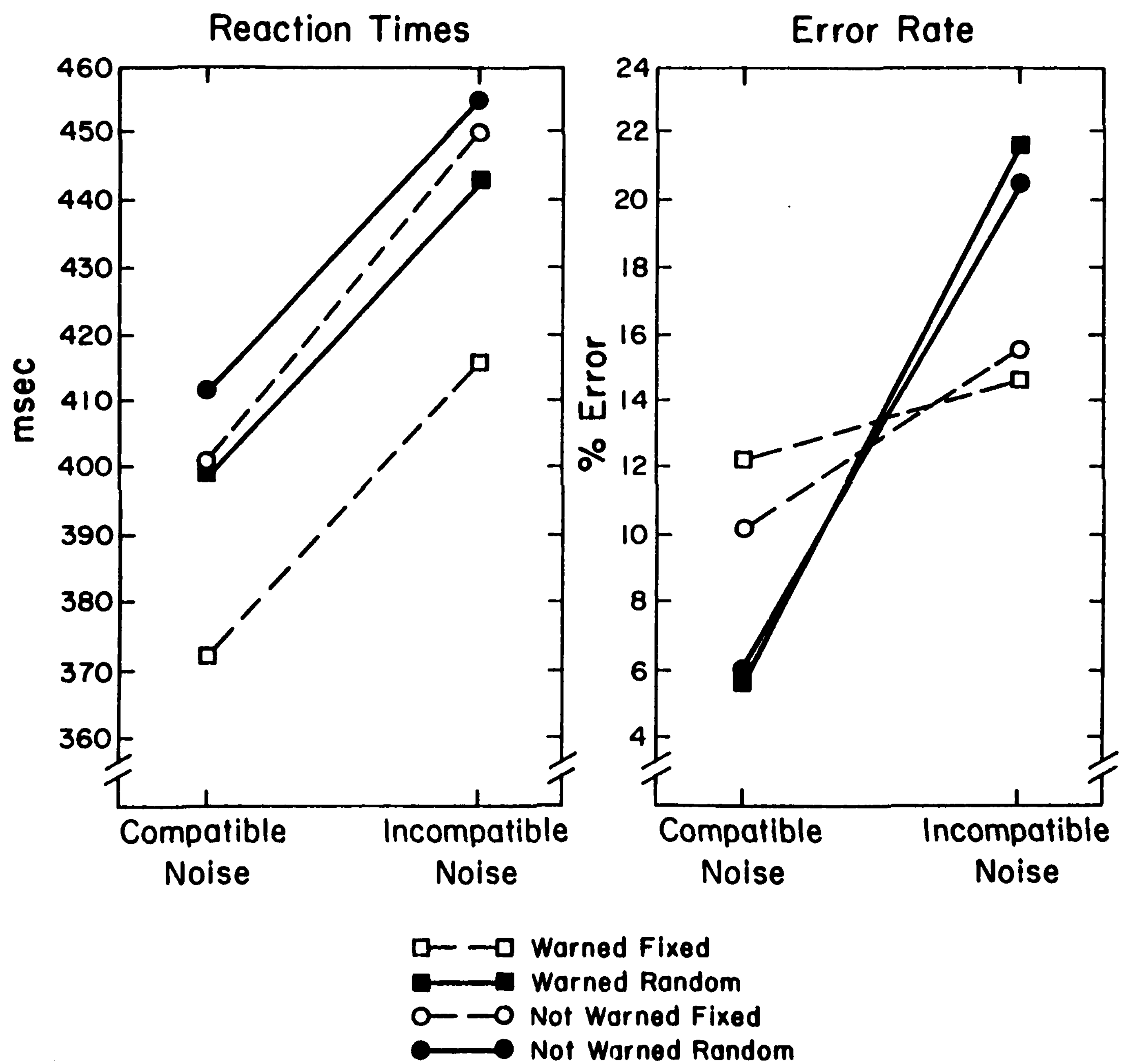

Reaction Times
Error Rate
msec
% Error
460
450
440
430
420
410
400
390
380
370
360
24
22
20
18
16
14
12
10
8
6
4
Compatible Noise
Incompatible Noise
Warned Fixed
Warned Random
Not Warned Fixed
Not Warned Random

FIG 1

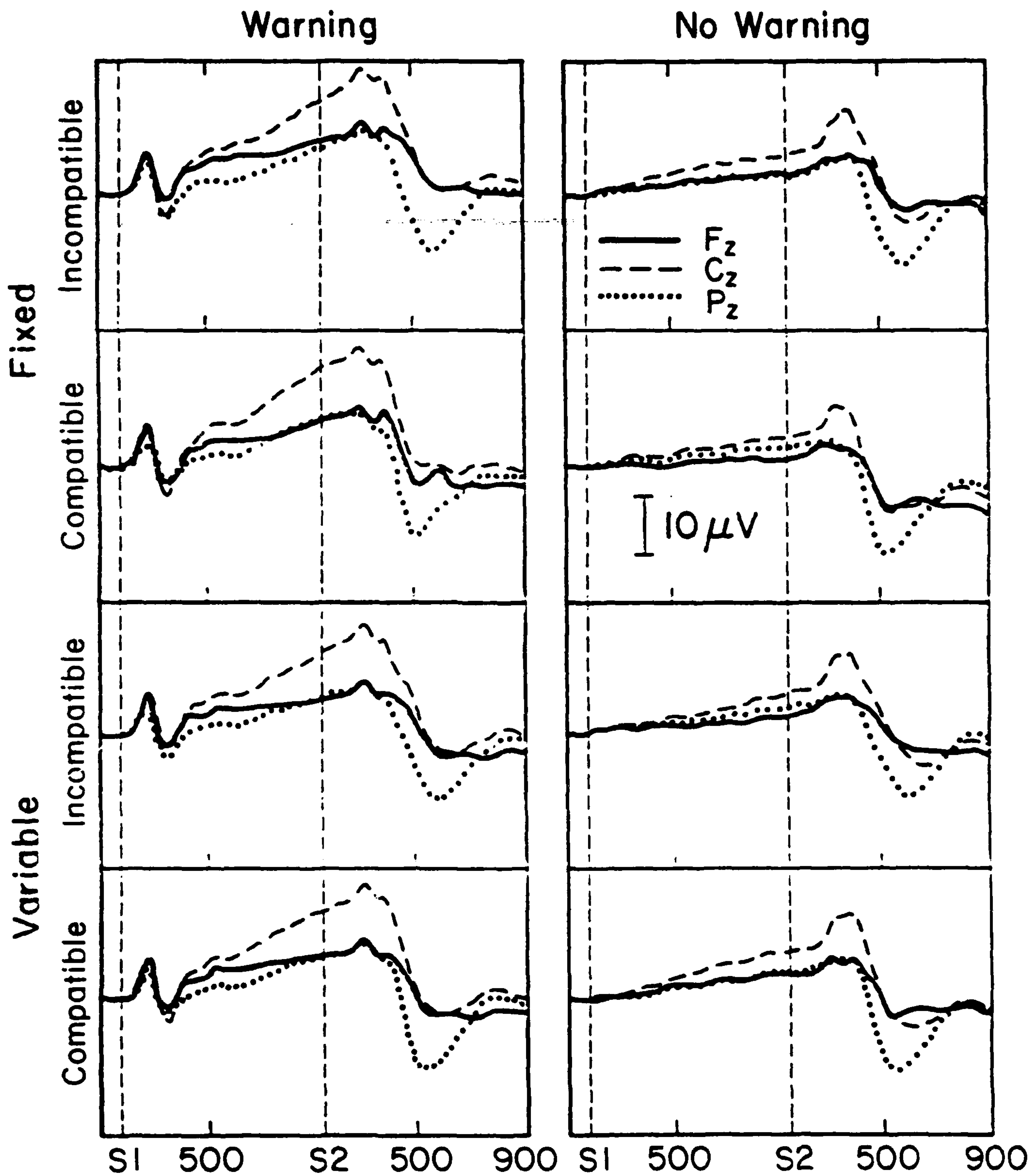

Warning
No Warning
Fixed
Variable
Incompatible
Compatible
Incompatible
Compatible
F_z
C_z
P_z
10 μV
S1
500
S2
500
900
S1
500
S2
500
900
msec

FIG 7

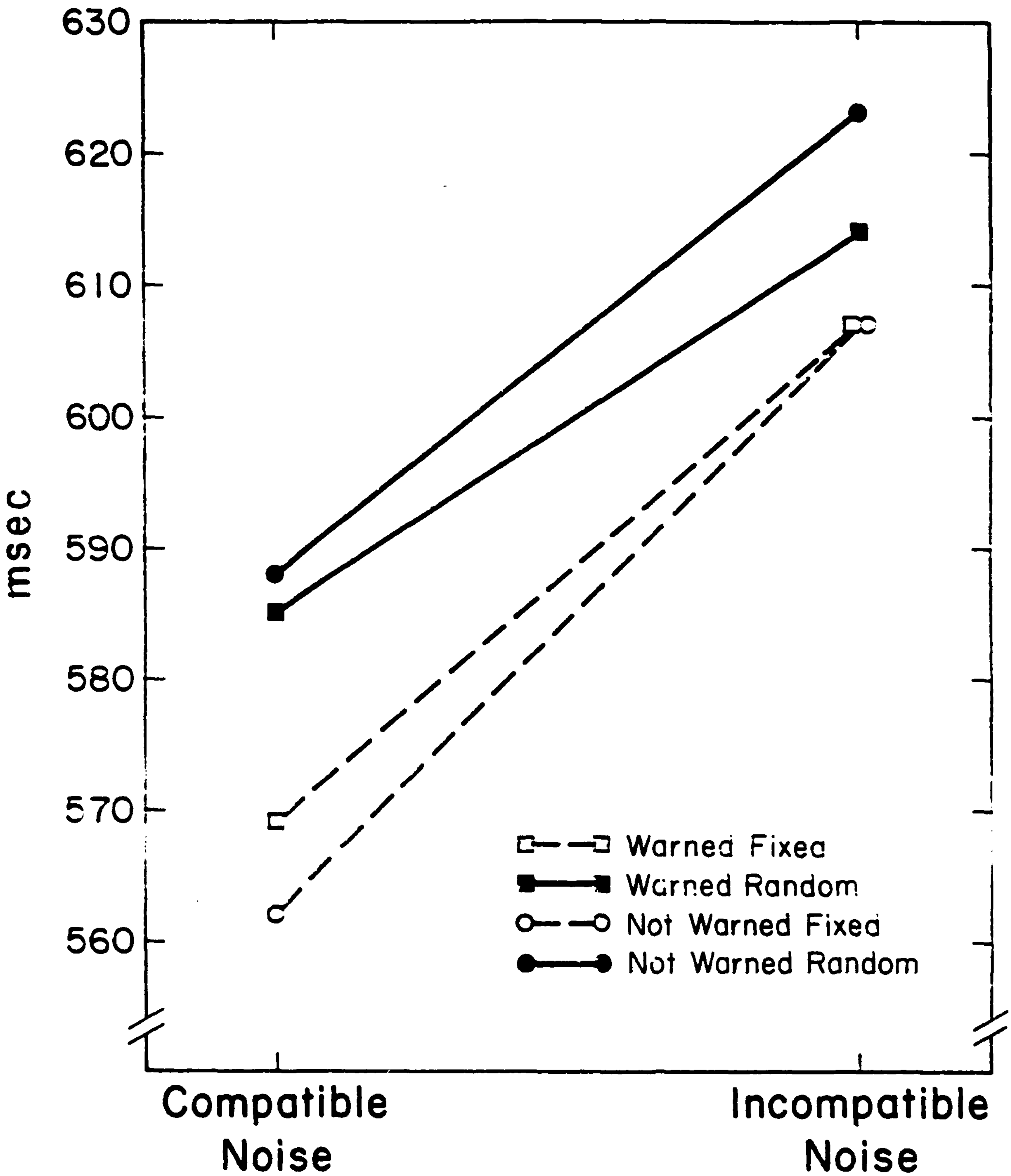
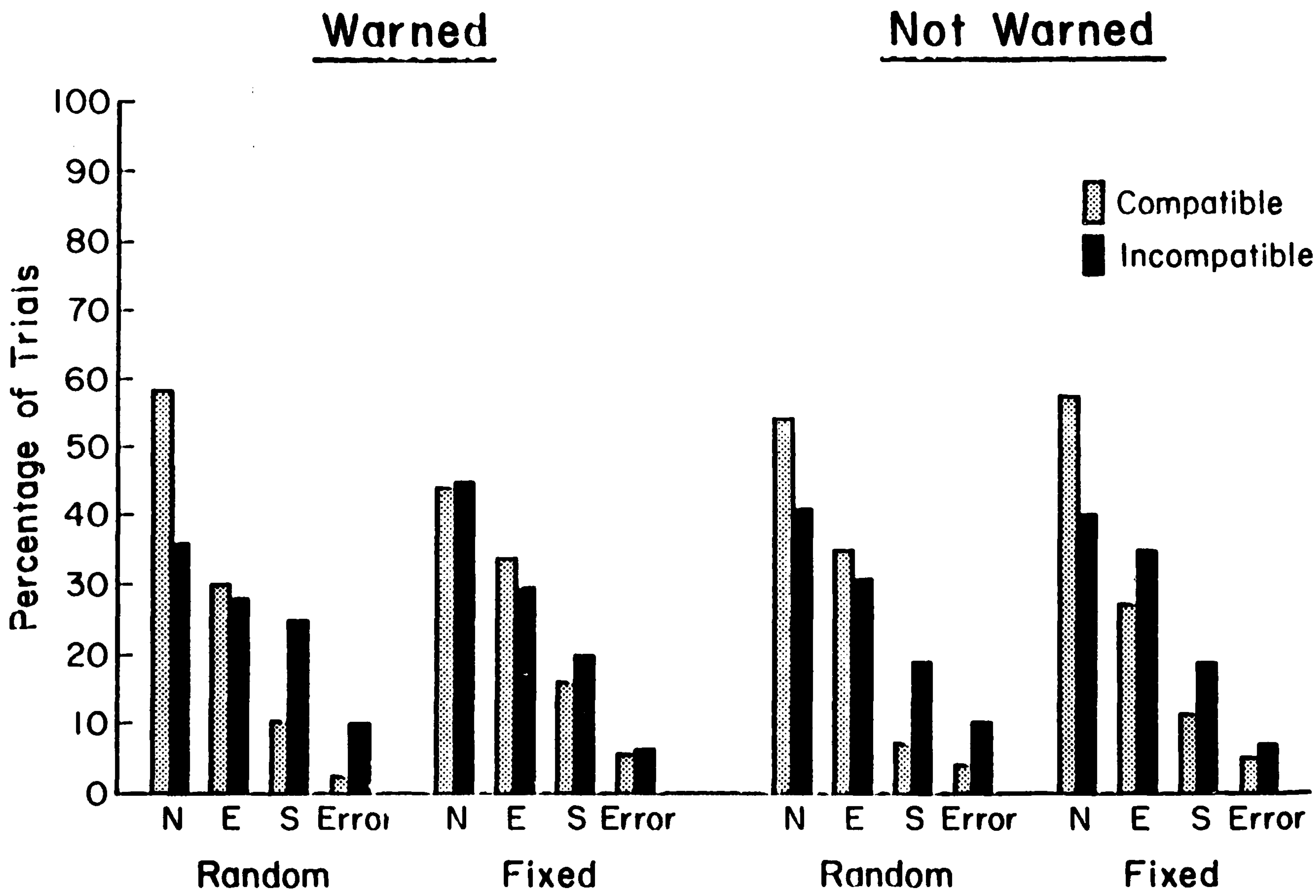
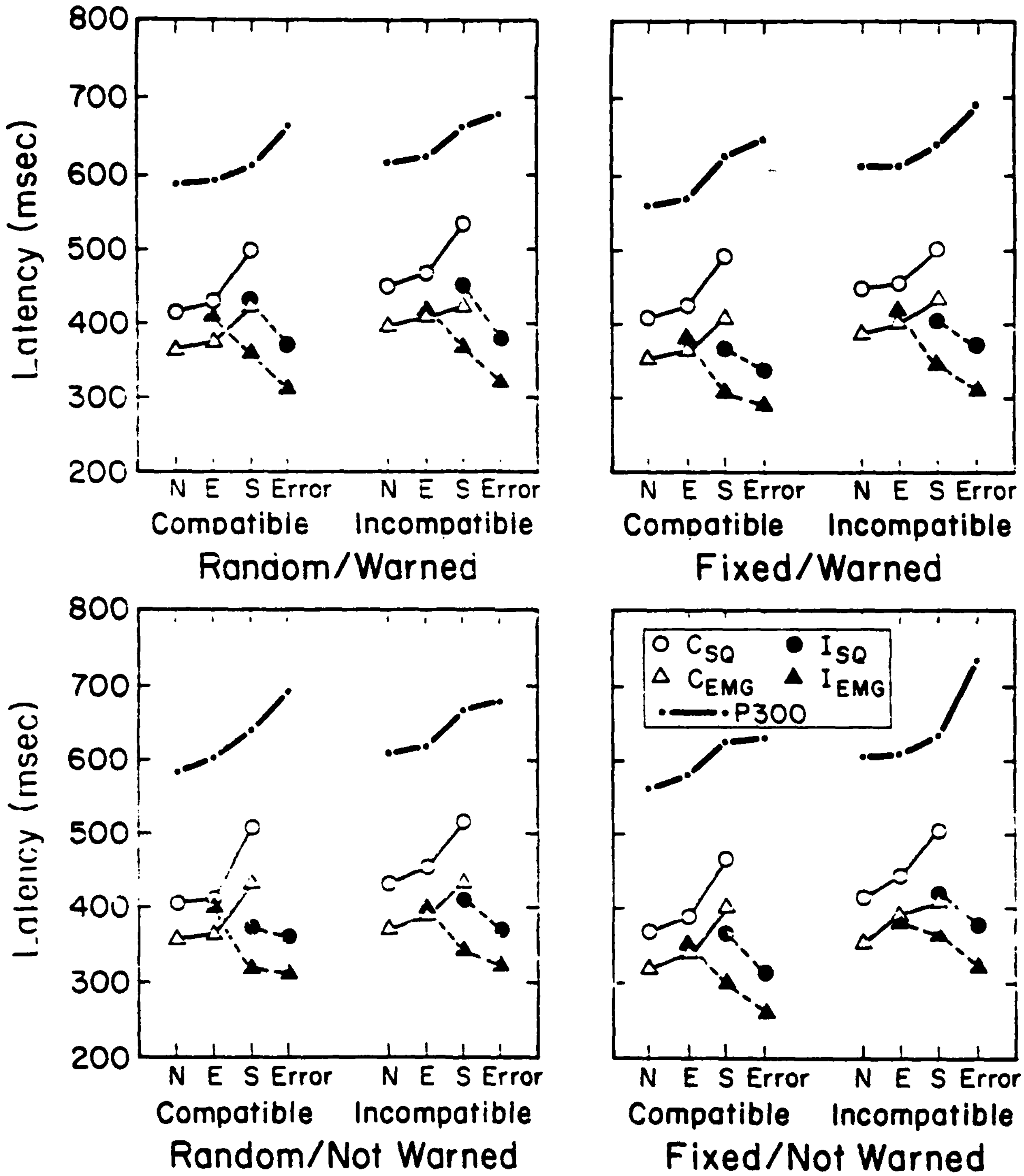

630
620
610
600
590
580
570
560
msec
Warned Fixed
Warned Random
Not Warned Fixed
Not Warned Random
Compatible Noise
Incompatible Noise

Fig 3



Warned
Not Warned
Percentage of Trials
100
90
80
70
60
50
40
30
20
10
0
Compatible
Incompatible
N
E
S
Error
Random
Fixed

FIG 4



Latency (msec)
800
700
600
500
400
300
200
N E S Error N E S Error
Compatible Incompatible
Random/Warned
Fixed/Warned
Random/Not Warned
Fixed/Not Warned
C_{SQ} I_{SQ}
C_{EMG} I_{EMG}
P300

Fig 5

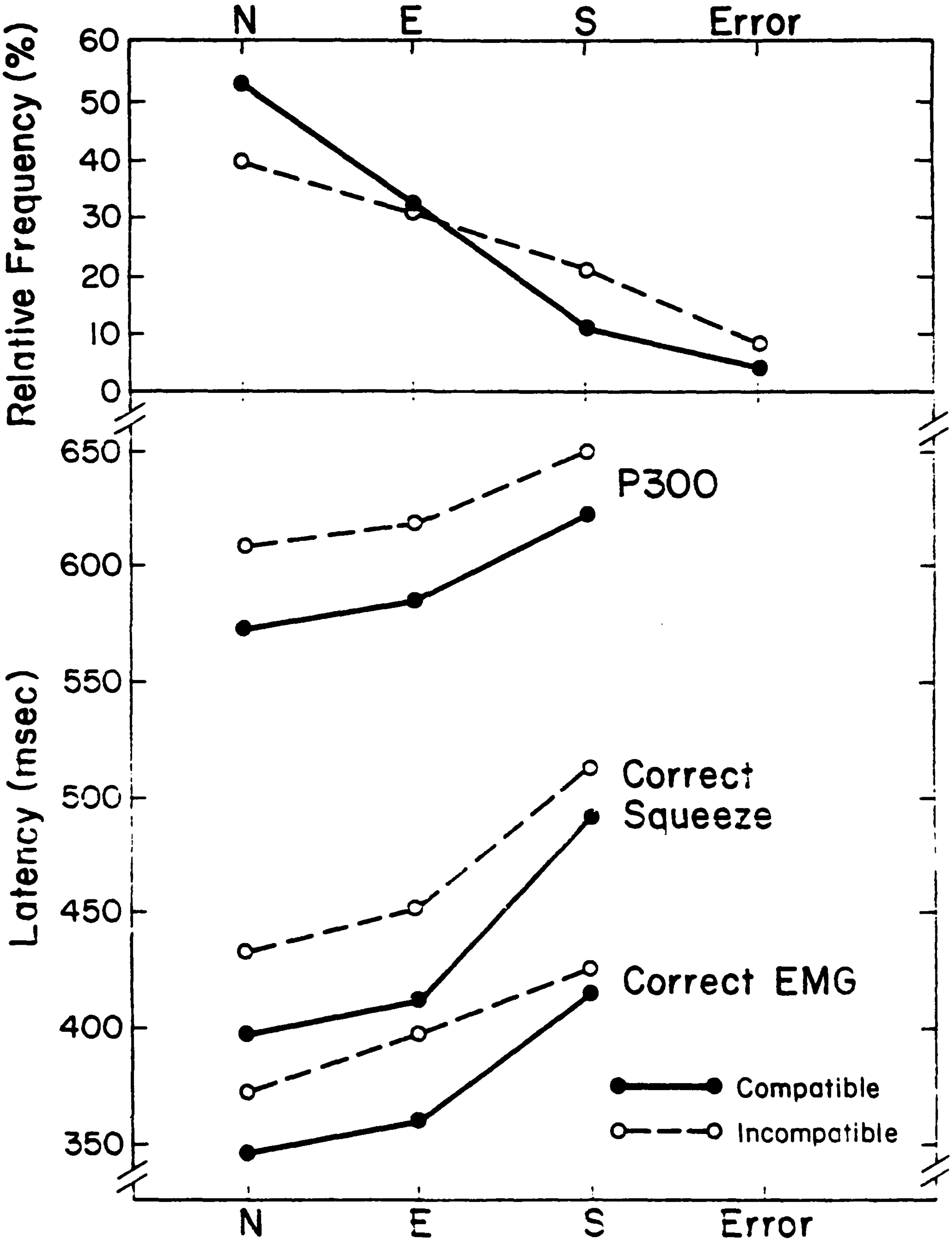
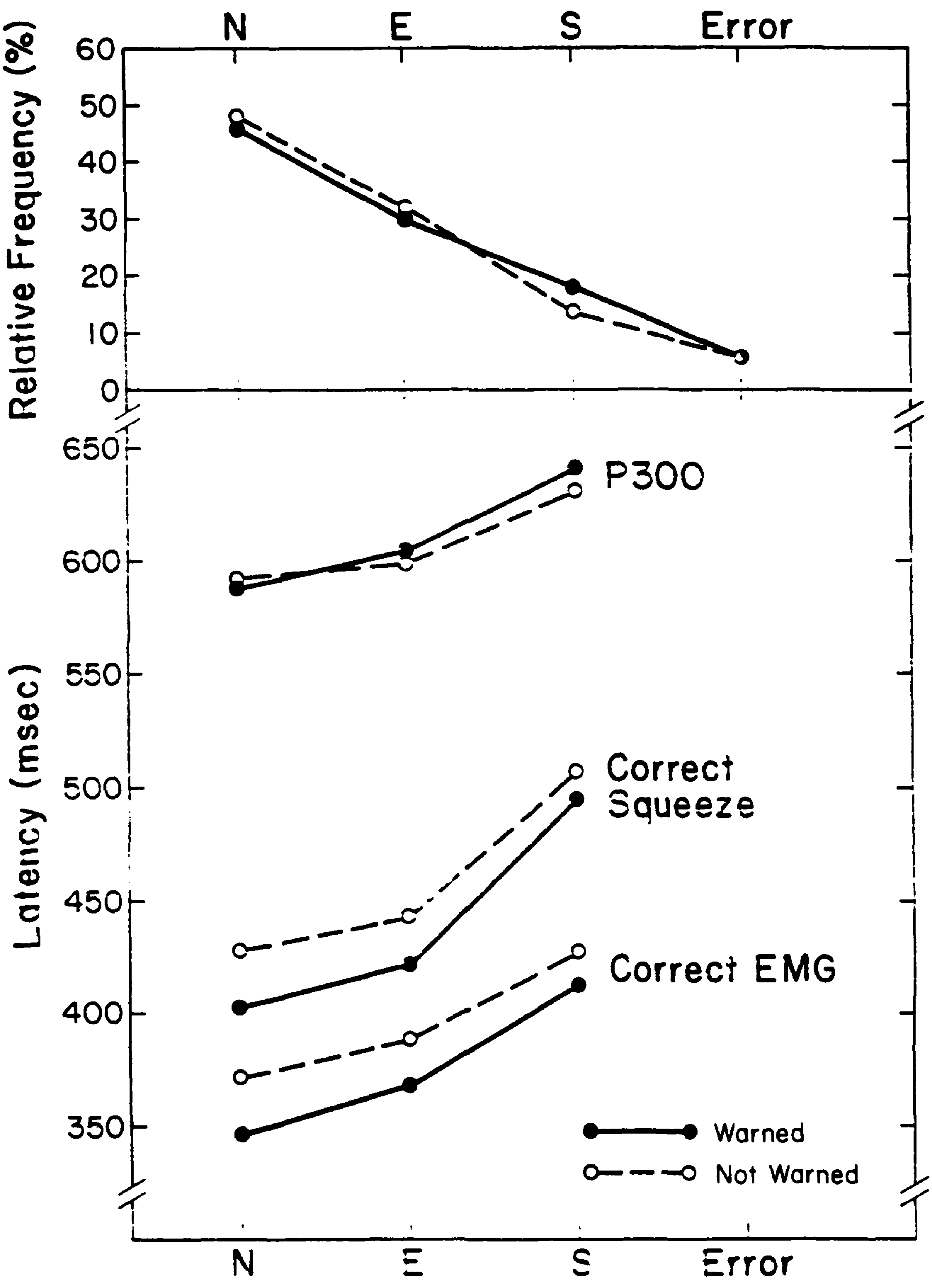
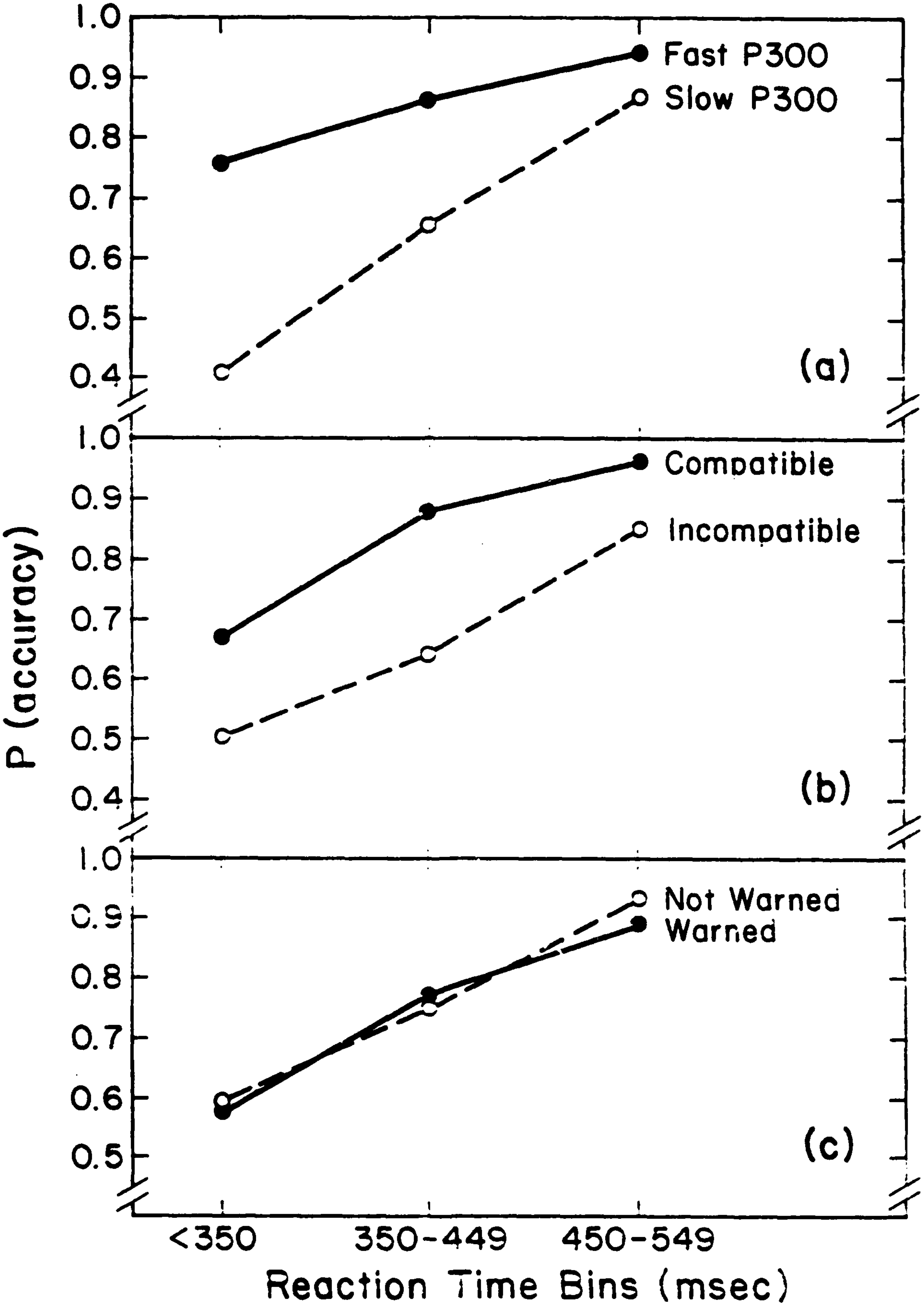

N
E
S
Error
Relative Frequency (%)
60
50
40
30
20
10
0
650
600
550
500
450
400
350
P300
Correct
Squeeze
Correct EMG
Compatible
Incompatible
Latency (msec)
N
E
S
Error

FIG 6



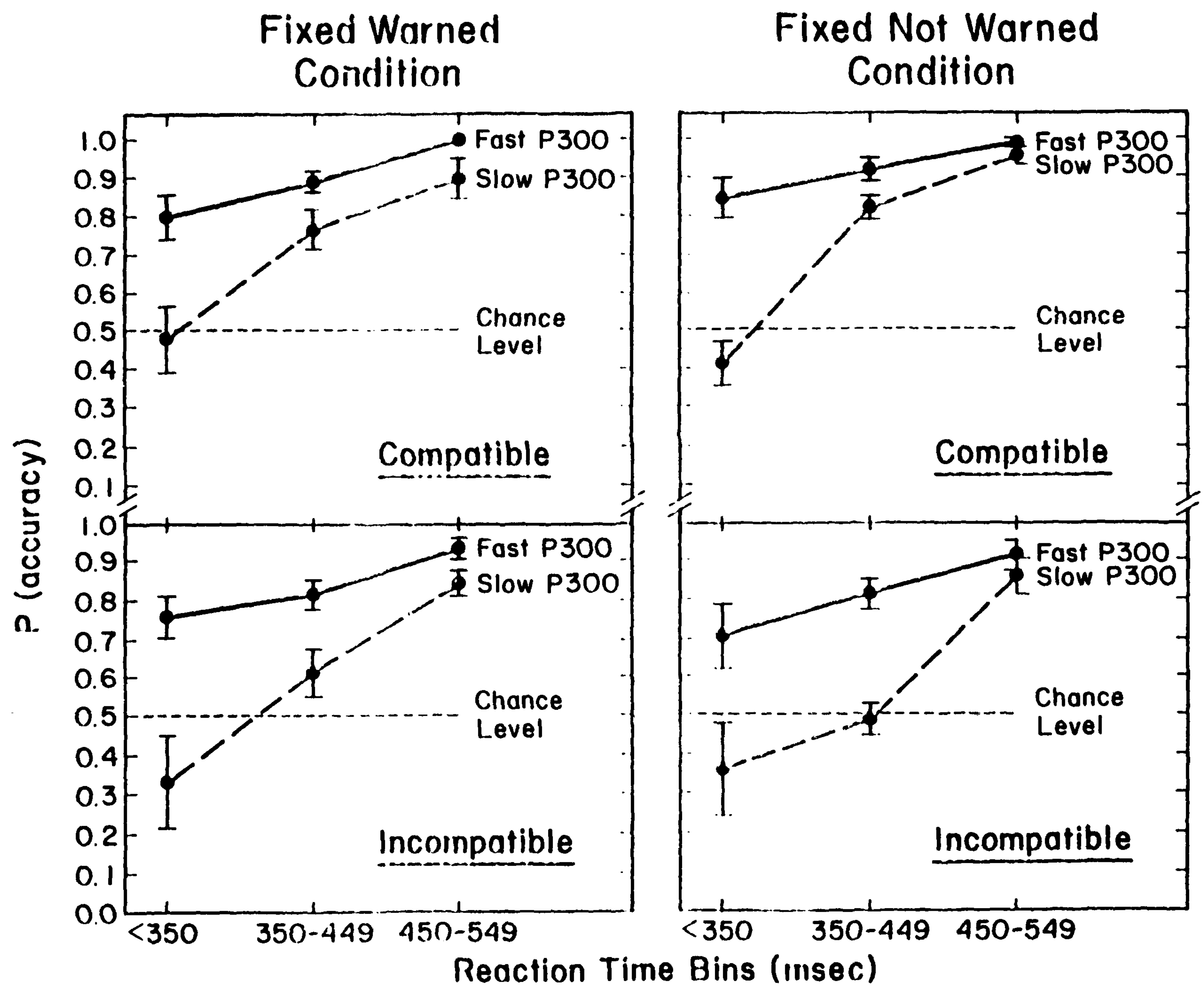
N
E
S
Error
Relative Frequency (%)
60
50
40
30
20
10
0
650
600
550
500
450
400
350
Latency (msec)
P300
Correct
Squeeze
Correct EMG
Warned
Not Warned
N
E
S
Error

Fig 7



1.0
0.9
0.8
0.7
0.6
0.5
0.4
Fast P300
Slow P300
(a)
Compatible
Incompatible
(b)
Not Warned
Warned
(c)
P (accuracy)
<350
350-449
450-549
Reaction Time Bins (msec)

Fig 8



Fixed Warned Condition
Fixed Not Warned Condition
P (accuracy)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Fast P300
Slow P300
Chance Level
Compatible
Incompatible
<350
350-449
450-549
Reaction Time Bins (msec)

Fig 9A

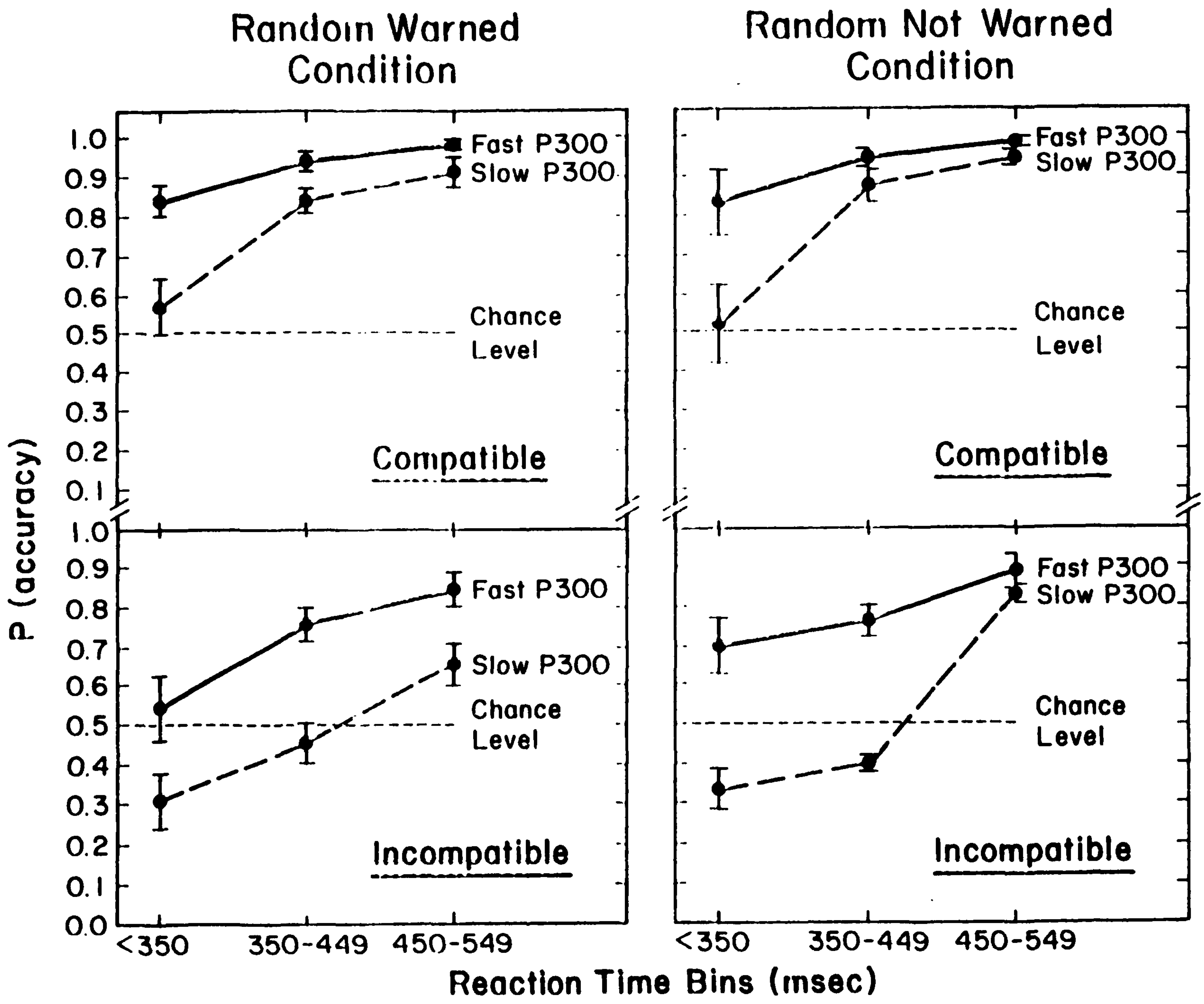

Random Warned Condition
Random Not Warned Condition
P (accuracy)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Fast P300
Slow P300
Chance Level
Compatible
Incompatible
<350
350-449
450-549
Reaction Time Bins (msec)

Fig 9B

A SIMULATION STUDY OF LATENCY MEASURES OF COMPONENTS OF EVENT-RELATED POTENTIALS

Gabriele Gratton, Arthur F. Kramer, and Michael G.H. Coles

Cognitive Psychophysiology Laboratory

Department of Psychology

University of Illinois at Urbana-Champaign

Running title: Simulation study of latency measures

ACKNOWLEDGEMENTS

The study presented in this paper was supported in part by a grant from the Air Force Office of Scientific Research, contract #F49620-83-0144, dr. Al Fregly Program Director. A partial report of some of the data was presented at the 24th Annual Meeting of the Society for Psychophysiological Research, Milwaukee, Wisconsin, October 18-21, 1984.

Introduction

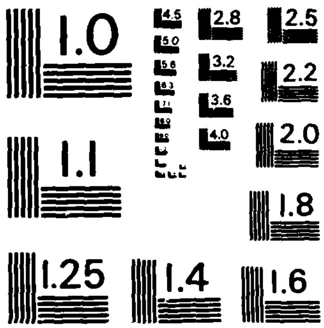
Recent trends in the study of the Event-Related Brain Potential (ERP) emphasize the advantage of the analysis of the ERP in terms of constituent components. Components are parts of the ERP that are interpretable as the manifestation of the activity of functional units, in response to, or in preparation for a particular event (Donchin, Coles, and Gratton, 1984). Components are customarily defined in terms of polarity, latency, scalp distribution, and sensitivity to particular experimental manipulations (Donchin, Ritter, and McCallum, 1978).

Theoretical speculation (see Donchin, 1981; Coles & Gratton, in press) and empirical evidence (see Pritchard, 1981; Duncan-Johnson, 1981; Duncan-Johnson and Donchin, 1982) indicate the psychological relevance of the latency of components of the ERP. However, since ERP components are estimated from data containing substantial amounts of noise, the measures of latency available only represent approximations to the "true" latency of the components. In this study, we compare the accuracy of estimates obtained with several different latency detection procedures. We also compare the impact of different signal-to-noise ratios on those estimates. In addition, this study provides guidelines for the choice of an appropriate procedure for the estimation of component latency. Of course, such guidelines are necessarily confined to the domain explored by this study.

A procedure intended to estimate ERP component latency should take into account the characteristics of both the signal (i.e., ERP component) and the noise (i.e., all other electrical activity recorded by the scalp electrode).

AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 6/9

UNCLASSIFIED 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

However, most component detection algorithms focus on the characteristics of the signal. We will show that, if characteristics of the noise are considered, the accuracy of the latency estimation can be improved.

The ERP noise consists of background EEG (random noise), and ERP components (systematic noise) which are active in the same time range as the component whose latency we intend to estimate (that is, the target component). A thorough evaluation of different latency estimation procedures can be obtained only when both these sources of noise are considered. However, most studies using simulated conditions only examine the effect of random noise (intended to simulate background EEG) on the accuracy of latency estimates (see Woody, 1967; Pfefferbaum, 1983).

We use "background EEG" as a label for the electrical brain activity which is not time-locked to the external event. It is generally assumed that the background EEG has a "random" phase in relation to the triggering event. However, this activity is not really random, in the sense that some frequency bands may be dominant. Furthermore, the presence of strong auto-correlation functions in the background EEG activity may affect signal detection procedures which are based on the autoregressive properties of the signal (cf. autocorrelation procedures). Therefore, to obtain veridical noise conditions, the frequency characteristics of the background EEG must be reproduced in the simulated waveforms. Unfortunately, the characteristics of the background EEG activity occurring during an ERP experiment are not well-known. Several studies have concurrently examined ERP and EEG frequency power spectra (for example, see McCarthy & Donchin, 1978). However, the presence of ERPs make the frequency power spectra obtained with such procedures poor estimates of the background EEG activity (see Ungar & Basar, 1976, for a related discussion on the utility of

"Wiener's filter", Wiener, 1949).

Several procedures have been adopted to simulate the background EEG noise. Two criteria for the choice of such procedures have been stressed: face validity of the noise, and ability to control and manipulate noise characteristics. Unfortunately, the characteristics of the background EEG noise are known only in part and are not necessarily constant over different experimental conditions. Therefore the simulation of background noise is difficult. An example of a first order autoregressive noise simulation procedure is illustrated in Donchin and Herning (1975).

We believe that we can obtain a good approximation to the background noise by using the deviation of single trial ERPs from the average ERP. This definition of noise corresponds to that adopted by the signal averaging technique. It may be considered valid in those cases in which the latency and amplitude of a component are reasonably constant over trials. However, when background noise is simulated in this way control over noise characteristics is sacrificed for face validity.

As we have noted, a second source of noise is provided by those components which are active in the same time segment as the target component. In this case, the average value of the noise over a large number of trials represents an estimate of the amplitude of the overlapping component(s). Note that this value is not equal to zero as is the case with averages of random noise. We may consider this kind of noise as "systematic". The error induced by "systematic" noise is particularly insidious because it may vary as a function of experimental manipulations. Furthermore, latency detection procedures might be differentially affected by overlapping components. To study the effects of overlapping components on the accuracy of latency estimates in the present study, we added to the

waveforms a set of components, in an attempt to simulate veridical conditions. The components varied in the degree of temporal and spatial overlap with the target component. The amplitude and latency of the components were systematically varied.

The relative amplitudes of the component (signal) and of the noise are important in determining the accuracy of the detection of the component. Several studies have demonstrated that detection accuracy increases monotonically with increases of the signal-to-noise ratio (Nahvi, Woody, Ungar, & Sharafat, 1975; Pfefferbaum. 1983; Wastell, 1977; Woody, 1967). However, different procedures may be differentially affected by the same increase in the signal-to-noise ratio, and a procedure which is more accurate at one signal-to-noise ratio may be less accurate at another signal-to-noise ratio (see a related discussion about Wiener's filter in Wastell, 1981).

Several studies have determined instances in which the amplitudes (see Squires, Wickens, Squires, & Donchin, 1976) and/or latencies (see Kutas, McCarthy and Donchin, 1977) of ERP components vary from trial to trial. Such variability suggests caution in the use of average waveforms to estimate component latency. In fact, a trial associated with a large component amplitude will have more weight in determining the average waveform than a trial associated with a small component amplitude. Therefore, the latency of the average waveform will be mostly determined by those trials associated with a large component amplitude. Latency variability may also affect amplitude estimates, by reducing the amplitude of the component peak (see Donchin & Heffley, 1979, for a discussion). In this case, an unbiassed estimate of component amplitude can be obtained by aligning each single trial on the component peak, and then computing the

average.

Such arguments clearly illustrate the advantages of a procedure which allows us to measure a component's latency from single trials rather than from averages. However, if we accept the assumptions of the averaging technique, the difference between single trials and averages can be conceptualized as a difference in the signal-to-noise ratio. In general, the signal-to-noise ratio in the averages should increase as a function of the square root of the number of trials entered in the computation (Coles, Gratton, Kramer, & Miller, in press).

In this paper we will consider two classes of techniques which are used to estimate the latency of ERP components: (a) preparation, or filtering, procedures, and (b) signal detection procedures. The former includes those filtering techniques which prepare the data for the actual latency estimation, carried out by the latter procedures. The purpose of these filtering techniques is to increase the signal-to-noise ratio, by amplifying the signal relative to the noise. We will consider two kinds of filtering techniques, frequency filters and spatial filters.

Frequency filters have been used in most analyses of ERPs. Their function is to eliminate electrical activity of undesirable frequencies. However, the effect of frequency filters on latency estimates has not been thoroughly explored. Both on-line and off-line filters are commonly used. On-line filters generally introduce phase shifts which result in distortions of the latency estimates. The magnitude of the phase distortion depends on the band-pass characteristics of the filter (see Duncan-Johnson and Donchin, 1977 for a discussion of the effect of high-pass filters on ERP waveforms). For this reason most researchers use broad band filters in the collection of ERP data. Off-line filters can be designed in such a way as to avoid the

introduction of phase shift. Such filters are often used in the preparation of the ERP data for signal detection techniques. In this study, we will focus on low-pass, off-line filters with no phase distortion (Ruchkin & Glaser, 1979).

Scalp distribution information has not often been used in the preparation of data for latency estimation (see Nahvi et al., 1975, for an attempt to use multichannel information for improving signal detection). In general, researchers have simply selected one electrode location to use for further analysis (we will label this procedure "channel selection"). One of the goals of this study is to evaluate the use of scalp distribution information to improve the detection of a component at different signal-to-noise ratios. Scalp distribution information can be used by adopting a new procedure (Vector filtering, VF), recently proposed by Gratton, Coles, and Donchin (1983b, 1984, and in preparation). Vector filtering is based on the assumption that scalp distribution is a defining characteristic of an ERP component. Therefore scalp distribution information can be employed to discriminate among ERP components, and between ERP components and noise. We will compare the accuracy of the latency estimates obtained from data prepared with Vector filtering and with channel selection. The latter technique can be thought of as a linear filter giving a weight of 1 to the selected channel, and a weight of 0 to other channels. Vector filter can be thought of as a linear filter where the weights for each channel are chosen in order to optimize the detection of the target component.

The task of a signal detection technique is to detect the signal under noisy conditions (for a review of signal detection techniques in ERP research, see Coles et al., in press). Two types of signal detection

techniques are commonly used by ERP researchers in the study of a component's latency: peak picking and cross-correlational techniques. The techniques differ in the way they define the signal. The peak picking technique identifies a component as a peak, or trough, in a certain time window. Note that only the point of the peak is used to estimate the parameters (amplitude and latency) of the component. Cross-correlation techniques (Derbyshire, Driessen, & Palmer, 1967; Palmer, Derbyshire, & Lee, 1966) define a component in terms of its waveshape. In fact, they define a component as that segment of waveform whose shape maximally "resembles" an externally defined segment of waveform, labelled the "template." "Resemblance" is assessed with a correlation or a covariance measure. Woody (1967) proposed a particular variant of such procedure, where the template is "adapted" to the average waveform, and several iterations are possible. The difference between peak picking and cross-correlation techniques may be conceptualized in terms of the number of data points considered for component detection. Peak picking techniques use only a single point, while cross-correlation techniques use a set of points (a segment of the waveform).

Wastell (1977) investigated the utility of the iteration procedure proposed by Woody (1967). He found that, if an appropriate template has been selected, the iterations proposed by Woody do not improve the correct detection of the signal. Furthermore, Pfefferbaum (1983) found that Woody's iterations may produce artifacts indistinguishable from real components. He concluded that cross-correlational techniques (and Woody filter in particular) produce the best results at relatively high signal-to-noise ratios. These studies suggest that the reliability of a signal detection procedure should be evaluated at several signal-to-noise ratios. However,

filtering procedures affect the signal-to-noise ratio. Therefore, the impact of these filtering procedures on different signal detection procedures must also be considered.

Method

There were three phases to the present study. First, a series of simulated waveforms were generated according to a particular model of ERPs. Second, a series of procedures were applied to these waveforms to obtain latency estimates. These procedures included various filtering and signal detection techniques. Finally, the accuracy of the latency estimations was computed, and the merits of each procedure were evaluated.

Model

The study was based on the following model:

$$E_{it} = a_i C_t + \sum_{j=1}^{n} (k_{ji} S_{jt}) + R_{it}$$

where:

E_{it} is the potential recorded at the electrode i at time t;

C_t is the amplitude of the target component at time t;

a_i is the weight of the target component at the electrode i;

S_{jt} is the amplitude of the overlapping component j at time t;

k_{ji} is the weight of the overlapping component j at the electrode i;

R_{it} is the background EEG noise at the electrode i at time t.

In adopting this model, we assume that each component is characterized by a particular scalp distribution, defined by a series of weights, one for each electrode location. This assumption is similar to that proposed for

the Vector filter procedure (Gratton, Coles, and Donchin, in preparation).

Single trials

Each assessment of the accuracy of the latency estimates was based on 100 repetitions, at three different electrode locations (labelled Fz, Cz, and Pz). Each repetition was called a "trial", and was obtained by adding a series of time vectors. The trials were constructed by adding a different noise vector to each of the identical 100 component vectors. The vectors consisted of 128 data points, that we considered as recorded at 100 Hz digitizing rate, starting 200 msec before a hypothetical stimulus. The average of the first 20 points was considered as an estimate of the "pre-stimulus" baseline level, and subtracted from the data.

Simulation of ERP components

One target and four non-target components were obtained by adding together five cosinusoidal waves. The amplitude, latency and duration (wavelength) of the cosinusoidal waves could be varied, and each component was simulated by using different parameters. The scalp distribution of each component was simulated by multiplying the vector by a different scaling factor for each electrode. The target component simulated the "P300" component, the non-target components simulated the "N100", "P200", "N200", and "Slow Wave" (SW) components. The parameters (amplitude, latency, duration, and scalp distribution) of each component do not correspond to data obtained in a particular experiment. However, an attempt was made to reproduce the parameters of the components described in the ERP literature (see Donchin et al. 1978). The parameters of N100 and P200 were not varied systematically, since they do not show temporal overlap with P300, and, therefore, should not affect the latency estimates. The amplitude and

latency of N200 and SW were systematically varied to simulate different degrees of overlapping with P300 that may be present in real data. A control condition in which only the P300 component was present was also included in the study. The parameters adopted for each component were the following (see Table 1).

Insert Table 1 about here

P300. P300 amplitude was varied systematically from 0 (absence of the component) to 500 units, with increments of 50 units. Each unit was intended to be equivalent to .1 microvolts, so that P300 amplitude varied from 0 to 50 microvolts. This manipulation of P300 amplitude allowed us to evaluate the different procedures over a wide range of signal amplitude conditions. Given the complex procedure we adopted to simulate noise (see below), in particular the presence of systematic noise, we could not express the true signal-to-noise ratio as an absolute value. However, we could compute the ratio between the amplitude of the signal and the root mean square amplitude (RMS) of the background EEG noise that was fixed at 100 (as shown later). We chose to label this value "signal-to-noise ratio". It varied systematically from 0 to 5, in half unit increments. The latency of P300 peak was set at 550 msec (post-stimulus), and the duration was fixed at 500 msec. A parietally maximum scalp distribution was simulated by assigning a weight of 1.2 to Pz, 0.8 to Cz, and 0.4 to Fz.

N100. The amplitude of N100 was fixed at 100 units (10 microvolts), the latency of the peak was 100 msec (post-stimulus) and the duration was 100 msec. A centrally maximum scalp distribution was simulated by assigning a weight of -1.2 to Cz, and -0.6 to Fz and Pz. Apart from the control

condition (in which N100 amplitude was set to 0), N100 parameters were fixed across all the experimental condition.

P200. The amplitude of P200 was 150 units (15 microvolts), the latency of the peak was 250 msec (post-stimulus) and the duration was 150 msec. A centro-frontal scalp distribution was simulated by assigning a weight of 1.0 to Cz and Fz, and 0.5 to Pz. As for N100, P200 parameters were fixed for all conditions, with the exception of the control condition, when P200 amplitude was 0.

N200. The amplitude and latency of N200 were independently manipulated, with two levels each. N200 amplitude was set at either 50 units (5 microvolts) or 100 units (10 microvolts), and the two peak latency levels were 300 and 400 msec (post-stimulus). These manipulations provided different degrees of overlap with P300, the overlapping being maximum when N200 amplitude was 100 units and the latency 400 msec, and minimum when the amplitude was 50 units and the latency 300 msec. The duration of N200 was fixed at 250 msec. A frontally maximum scalp distribution was obtained by assigning a weight of -1.2 to Fz, -0.8 to Cz, and -0.4 to Pz. Given this set of weights, N200 scalp distribution was clearly different from that of P300.

Slow Wave. As for N200, Slow Wave amplitude and latency were independently manipulated with two levels each. The two levels of Slow Wave amplitude were 100 units (10 microvolts) and 200 units (20 microvolts), and two latency levels were 800 and 1080 msec (post-stimulus). Slow Wave duration was fixed at 800 msec. A parietally positive and frontally negative scalp distribution was obtained by assigning a weight of +0.4 to Pz, 0.0 to Cz, and -0.4 to Fz. Note that this set of weights provided a pattern of scalp distribution close to that of P300.

The complex ERP. The five components described above were added to obtain complex ERP waveforms. An example of complex ERP waveform is presented in the lower part of Figure 1.

Insert Figure 1 About Here

The manipulation of the components' amplitudes and latencies were factorially combined. Thus the design included 2x2 (amplitude and latency) P300 manipulations, 2x2 Slow Wave manipulations, and four repetitions of the control condition, with a total of 20 conditions of component overlap. As there were 11 P300 amplitude levels, 220 basic ERP waveforms were obtained for each electrode. The design of the study will be presented later in a more detailed manner.

Background EEG simulation

Background EEG activity was simulated by obtaining non-event related activity from a set of 100 trials recorded from an individual subject in an oddball experiment (Fabiani, Gratton, Karis, and Donchin, in preparation). In this experiment, the subject was presented with one of two tones on any given trial. The tone probability were .2 and .8. ERPs were recorded at Fz, Cz, and Pz. The on-line filtering procedure included a low-pass filter with a half amplitude cut-off point at 35 Hz, and a high-pass filter with a time-constant of 8 sec. Vertical EOG was recorded from above and below the right eye, and ocular artifacts were corrected with a procedure described in Gratton, Coles, and Donchin (1983a). Separate averages were obtained for frequent and rare trials and for each electrode. "Non-event related" activity was obtained by subtracting the average appropriate for each event

from each single trial record. The average spectral power density functions of the non-event related activity for each electrode is shown in Fig. 2a.

Insert Figure 2 About Here

As evident from Fig. 2a, the power spectra of the "non-event related" activity do not show much activity in the alpha frequency band (8-12 Hz).

This procedure yielded a set of waveforms whose average is a flat line. However, the variability from trial to trial is not equal for all time points and electrodes. In particular, larger intertrial variance was observed at a latency of approximately 300 msec, and smaller variance during the prestimulus period. Under these conditions, the characteristics of these waveforms could not be considered "stationary" over the whole epoch, and, therefore they could not be considered as good estimates of the background EEG noise. A further disadvantage was that the impact of noise could vary as a function of the latency.

Therefore we chose to standardize each timepoint (and electrode), with a mean of 0, and a standard deviation of 100 units (10 microvolts). We considered the resulting 100 waveforms for each electrode as our "simulated" background EEG activity. The relative average power spectra density functions for each electrode are shown in Fig. 2b. As shown in this figure, these power spectra did not differ significantly from those obtained before the standardization process. However, as a control condition, we replicated some of the procedures with "non-standardized" waveforms.

An example of single trial waveforms obtained by adding the complex ERP waveform and the standardized background noise is shown in the upper part of Figure 1.

adjustment procedure itself.

Overlapping component conditions. Logarithms of MSE (and number of trials required to obtain an error of 3 msec) for four different component overlap condition with two spatial filtering and two signal detection procedures are shown in Figure 7.

Insert Figure 7 About Here

The "no-component overlap" condition is shown in the upper left panel for a comparison. The other three conditions shown in the figure were "small component overlap" (N200 amplitude = 50 units, N200 latency = 300 msec, Slow Wave amplitude = 100 units, Slow Wave latency = 1280 msec), "large N200 overlap" (N200 amplitude = 100 units, N200 latency = 400 msec, Slow Wave amplitude = 100 units, Slow Wave latency = 1280 msec), and "large Slow Wave overlap" (N200 amplitude = 100 units, N200 latency = 300 msec, Slow Wave amplitude = 200 units, Slow Wave latency = 1000 msec). Frequency filtering was not applied to the data shown in Figure 7. An inspection of this figure reveals that the component overlap impaired the accuracy obtained with each procedure to a different degree. In particular, procedures based on channel selection (Pz) were markedly affected by component overlap, both of N200 and Slow Wave. Procedures based on Vector filtered data were also affected by Slow Wave overlap, but were not affected by N200 overlap. We should note here that the scalp distribution of Slow Wave was rather close to that of P300, while the scalp distribution of N200 was very different. Thus, estimates obtained on Vector filtered data are not affected by overlapping components with a scalp distribution very different from that of P300, but are affected by an overlapping component with a scalp distribution similar

point in the epoch). At higher signal-to-noise ratios, central values becomes progressively more represented. The mode tends generally to correspond to the actual P300 latency. An exception to this general rule can be observed at a signal-to-noise ratio of 2.5 for the peak-picking algorithm on Pz waveforms. A skewed distribution indicates the presence of systematic error (the average estimated latency does not correspond with the real P300 latency).

Latency adjusted average waveforms obtained with cross-correlation and second and third iterations of Woody filter for the "non-overlapping" component condition, at extreme levels of the signal-to-noise ratio, and no frequency filter, are shown in Figure 6.

Insert Figure 6 About Here

Inspection of this figure reveals that, even when no ERP component is present (signal-to-noise ratio is equal to 0), the latency adjustment procedure "creates" one. When the component is large, the distortion produced by the latency adjustment is negligible. This finding is in agreement with the results reported by Pfefferbaum (1983). The artifactual component created by the latency adjustment appears larger when the signal detection algorithm is applied to Pz waveforms, than for waveforms obtained with Vector filter. This effect is confounded in part with an overall reduction in amplitude produced by Vector filter. The artifactual component appears also to have the same amplitude if the latency adjustment is obtained after the cross-correlation procedure, Woody filter with one iteration, or Woody filter with two iterations. Thus, the problem does not seems to be related to the number of iterations, but rather to the latency

high as 50%, at very high signal-to-noise ratios, and when no frequency filter is applied. Third, the use of Vector filter as a spatial filtering technique reduced the error in latency estimation in comparison with channel selection (Pz). This advantage is evident at middle and high signal-to-noise ratios. At low or middle signal-to-noise ratios (.5 to 2.0) the advantage of Vector filter is comparable to that of cross-correlation. However, the advantage of Vector filter rarely reaches the 50% level, and is usually about 25%. The advantages of Vector filter and of cross-correlation appear to be independent. Fourth, low-pass frequency filters produced marked improvements of the accuracy of latency estimation. The largest improvement was obtained with a 130 msec moving average iterated twice. Filters with a wider bandpass produced smaller improvement. However, this effect was particularly evident when a peak-picking algorithm was used for signal detection. The gain for cross-correlation was small. In fact, the effect of the frequency filters was to bring peak-picking to the same level of accuracy as cross-correlation. The gain obtained with Vector filter was unaffected, and in fact the smallest MSEs were obtained by the joint use of frequency filters, Vector filters, and cross-correlation.

Histograms of the latency estimates for each single trial in the "non-overlapping component" condition, and no frequency filter, are shown in Figure 5.

Insert Figure 5 About Here

The distribution at a signal-to-noise ratio of 0 was approximately rectangular, indicating that no point was more likely to be chosen than any other when no signal was present (apart for a small preference for the f -st

No-overlapping component condition. The basic design was devised to permit a comparison of the accuracy of several procedures over a wide variety of signal and noise conditions. As a reference point we will first present the data obtained in the condition in which no overlapping component was present. The MSE values (averaged across 400 repetitions) for this condition are shown in Fig. 4.

Insert Figure 4 About Here

As a reminder, in this and most of the following figures the abscissa represents the signal-to-noise ratio (or, P300 amplitude), while the ordinate represents the MSE (logarithmic transform). Note that the results obtained with the second and third iterations of the Woody filter are not shown in this figure. The MSE values obtained with Woody filter were very close to those obtained with cross-correlation.

Several important effects are apparent in figure 4. First, variations of the signal-to-noise ratio produced the largest effects on the accuracy of estimation. At a signal-to-noise ratio of 0 all the procedures gave about the same results. The MSE at this signal-to-noise ratio is close to that which would be obtained by picking points at random in the temporal window. In fact, the log MSE obtained in this way is 2.2. By increasing the signal-to-noise ratio, exponential decreases of the MSE can be observed (the functions approximate a line in the figure because of the logarithmic scale used for the ordinate). At a signal-to-noise ratio of 5, the MSE is one tenth of that found at a signal-to-noise ratio of 0. Second, the use of cross-correlation as signal detection procedure yielded lower MSE than peak-picking. The gain in accuracy obtained with cross-correlation may be as

$$N = (MSE / 3)^2 + 1$$

where N is the number of trials required to obtain a standard error of 3 msec. Note that this value is only an approximation. In fact, it requires that (a) the distribution of the single trial estimates is normal distribution, and (b) that the sample mean is not systematically different from 550 msec. The first assumption is violated, since only values inside the time window (300 to 800 msec) are possible. However, the distribution of the single trial estimates is approximately normal when the signal-to-noise ratio is larger than 1. Examples of distributions of single trial estimates for different signal-to-noise ratios will be shown later. The second assumption may also be violated in some cases, but it holds in most cases. Since the number of trials required to obtain a standard error of estimate of 3 msec are related to the MSE, we simply added a scale reporting the corresponding values for this dependent variables in most of the figures in which MSE (or log-MSE) is used.

RESULTS AND DISCUSSION

The result section will be divided into two parts: first, we will present the results obtained from the basic design, second, we will discuss a series of additional analyses we ran to investigate the effect of "non-standardized" background noise, variations of P300 duration, and variations of the parameters of Vector filter, on the accuracy of the latency estimates.

Basic design

different procedures to be proportional to the absolute value. With a logarithmic scale, similar percent differences at different absolute levels of MSE will be represented equally. For example, a difference of 20% at an absolute level of MSE of 20 ms will be represented equally to a difference of 20% at an absolute level of MSE of 200 ms. This would not be the case with a linear scale.

Another dependent variable we used was an approximate estimate of the number of trials required to reduce the standard error of estimate to 3 msec. This measure was intended to provide an estimate of the relative power of the different procedures, and was obtained as follows. The MSE can be considered an estimate of the standard deviation of the population of single trial P300 latency estimates for each condition (note that the population mean is known). However, the mean of the single trial estimates of the sample may not correspond to the mean of the population (550 msec). If the normality assumption is met, we can compute the theoretical distribution of the population of sample means from which the mean of our sample is extracted. Following the theorem of central tendency, this distribution will have a width (measured by the standard error of estimate) proportional to the MSE (standard deviation) and inversely related to square root of the number of trials used to compute the mean. By increasing the number of trials we may reduce the standard error of estimate to any desired value. Thus, by appropriately setting the sample size, we may in theory obtain that the standard error of estimate is equal to 3 msec, whatever is the value of the MSE. In fact, we can compute the sample size required with the following equation:

noise components were added to each condition). A list of the experimental conditions for the basic design is given in Table 2.

Insert Table 2 about here

Error (accuracy) estimation

As mentioned above, 100 repetitions were obtained for each condition. To assess the accuracy of latency estimation obtained with each procedure under each condition, the mean square error (MSE) value was considered. This value was obtained as follows:

$$MSE = \sqrt{\mathrm{Sum}_{i=1}^{n} (l_i - L)^2 / n}$$

where:

MSE is the mean square error of latency estimate;

n is the number of trials;

l_i is the latency estimate at trial i;

L is the P300 peak latency (550 msec).

Most of the figures presented later in this paper show variations of the MSE value (or of its logarithmic transformation) as a function of variations of the signal-to-noise ratio. The method adopted to estimate the signal-to-noise ratio is explained earlier in this paper. As a reminder, the signal-to-noise ratio was proportional to the amplitude of P300. In the most of plots presented hereafter, a logarithmic scale is used. This scale was chosen because we assumed the variability in the MSE obtained with

post-stimulus and ended 800 msec post-stimulus (respectively, 250 msec before and 250 msec after the P300 peak). The duration of the template used for the cross-correlation algorithm was 500 msec.

Experimental design

It might be helpful to note the differences between the basic experimental design and the control conditions we ran to investigate particular problems. The basic experimental design consisted of a factorial manipulation of the simulated conditions and the analysis procedures described above. The simulated conditions yielded 220 different conditions (11 P300 conditions x 20 overlapping component conditions). For each condition we obtained 100 repetitions by adding simulated background EEG noise, sampled at random and with reselection from the 100 noise trials described above (note that relationship between waveform and recording electrode - i.e., Fz, Cz, and Pz was maintained). This yielded 22,000 waveforms. Each of these waveforms was filtered using the five filtering conditions. This produced a total of 110,000 waveforms for electrode, each of which was finally filtered spatially using either channel selection or Vector filter. Thus, we applied our four signal detection algorithms (peak-picking, cross-correlation, Woody with 2 iterations, and Woody with 3 iterations) were then used to obtain latency estimates for each of the 220,000 waveforms.

Note that the comparisons between signal detection algorithms and spatial filtering techniques (as well as their interaction) were based on repeated measures, while the comparison between frequency filtering procedures was based on independent measures, resulting in a nested design. Note that the component manipulation yielded independent measures (different

one data point and a template, which represents the signal to be detected. The segment of the ERP associated with the largest correlation is considered to contain the component (signal). Woody (1967) proposed the use of an adaptive template, obtained with an iterative procedure (Woody filter). The template for the first iteration is usually the average ERP, but sometimes arbitrary templates are used, like a sinusoidal wave (Pfefferbaum, 1983). However, in later iterations, the template is always extracted from the data by averaging the segments of ERP with a maximum correlation with a template at a previous iteration.

In this study we adopted a procedure which allowed us to evaluate both cross-correlation and Woody filter. We used a template equivalent to the P300 component we entered in the simulated waveforms as the first iteration of the Woody filter procedure. Thus, the first iteration corresponded to a cross-correlation algorithm, while the following iterations corresponded to successive iterations of the Woody filter. The estimate of P300 latency was obtained by selecting the central value of the ERP segment with the maximum correlation with the template.

The template we used for the cross-correlation technique was the target component itself. Therefore, the detection of P300 obtained with this algorithm may be more accurate than that which could be obtained in real (non-simulated) conditions, when the actual P300 waveshape is not perfectly known. To investigate the merit of different signal detection algorithms in cases in which we do not have a good representation of the P300 waveshape, we ran a control condition in which the duration of P300 was parametrically varied between 100 and 1000 msec, while the duration of the cosinusoidal wave adopted as template was fixed at 500 msec.

For each signal detection procedure, the temporal window began 300 msec

Insert Figure 3 About Here

As parameters for our Vector filter we chose a polarity angle of 15 degrees and an orientation angle of 300 degrees. These values do not correspond to the scalp distribution of P300. The relative scalp distribution associated with an orientation angle of 300 degrees is also shown in Figure 3, and corresponds to a parietal maximum, but central minimum scalp distribution. The rationale for this "paradoxical" choice is given above: we filtered for a scalp distribution which dissociated the parietal from the central electrode in an attempt to selectively reduce the noise contribution, and thus to increase the signal-to-noise ratio. The choice of the specific parameters was based on a previous study on real data in which they were found to produce on optimal discrimination between two groups of trials (rare and frequent trials in an oddball paradigm) with different P300 amplitude. As a control for our choice, we ran a condition in which we parametrically varied the parameters of the Vector filter (polarity and orientation), to determine which parameters resulted in the best improvement in the estimation of P300 latency.

Signal detection techniques

Two types of signal detection algorithms were used: (1) peak-picking, and (2) cross-correlation. The peak-picking algorithm is based on the detection of the maximum value in a prespecified temporal window. The cross-correlation algorithm involves the computation of a series of correlations between segments of the ERP waveform progressively shifted by

differences among electrodes. The 0 value expresses the condition in which the mean of the electrodes is equal to 0 (if the values of all electrodes are equal to 0, a polarity of 0 will be arbitrarily chosen).

b. An "orientation" angle describing the relative distribution of the potentials. In the three-electrode case, the deviations of each electrode from their mean can be plotted on a plane as three non-orthogonal axes (the angles between the axes will be equal to 120 degrees). This plane corresponds to a description of the variance across electrodes (Gratton et al, in preparation). Any pattern of deviations can be described by an axis in this plane. We label "orientation" the angle between this axis and an arbitrary reference axis. Thus, any pattern of relative distribution of the event-related potentials will be associated with an orientation angle. A reference axis which corresponded to the relative distribution of electrode values was used for our simulated P300 (maximum at Pz, minimum at Fz, with the Cz exactly a half distance between the two). Any other relative scalp distribution could be described by an "orientation" angle with this axis. The "orientation" angles for the other components we used in this study were:

a. 270 degrees for the N100;

b. 135 degrees for the P200;

c. 0 degrees for the N200;

d. 0 degrees for the Slow Wave.

A graphic representation of these angles and the associated scalp distributions is presented in Figure 3. A procedure to obtain orientation values for any scalp distribution is described in Gratton et al. (in preparation).

consequently an improvement of the signal-to-noise ratio, will be obtained by filtering for a scalp distribution where noise activity will not be strongly represented. In particular, filtering for a scalp distribution which dissociates the activity of electrodes with strong noise coherence functions is advisable. This rationale was the basis of our choice of Vector filter parameters. In fact, these parameters produced a dissociation between the activity of the parietal and central electrodes.

As mentioned above, the parameters of Vector filter are a series of angles (polar notation) which identify a particular scalp distribution. These angles reflect the weight given to each electrode in the filtering procedure. However, this method of describing a Vector filter is impractical because (a) one of these angles is redundant; (b) recovery of information about the overall polarity of the distribution is difficult. Hence, we (Gratton et al., in preparation) proposed a different way of describing a scalp distribution. This description is based on the separation of the information related to the common trend across electrode sites (polarity) and to the relative patterning of the electrodes (orientation). In the three-electrode, case this approach allows us to describe a scalp distribution (and therefore a Vector filter) by means of two angles:

a. A "polarity" angle, reflecting the relative value of the mean of the electrodes in comparison with their variance. A negative value of the polarity angle indicates an overall negative scalp distribution, a positive value, an overall positive scalp distribution. The polarity angle may assume values between +90 and -90 degrees. The extreme values describe scalp distribution characterized by a positive or negative mean value, but no difference among electrode sites. Intermediate values indicate larger

each data point as a vector (data vector), in a space defined by the electrode locations (see Gratton et al., in preparation). Particular patterns of scalp distribution correspond to axes in this space and may be identified by a series of angles with the electrode axes. These angles express the weight of the scalp distribution on each electrode. A basic assumption of Vector filter is that each component is defined by a specific scalp distribution. The presence of the component at each data point may be assessed by projecting the data vector onto the axis corresponding to the scalp distribution of the component.

The result of the vector filtering operation is an estimate of the amplitude of the target component (defined in terms of scalp distribution) for each datapoint. However, rather than considering this procedure as a "signal extraction" technique, we prefer to consider it as a filtering technique used to prepare the data for signal detection. By applying Vector filter, we eliminate from our data the part of the electrical activity that, because of its scalp distribution, does not represent the target component. In fact, we may consider it as a way of increasing the signal-to-noise ratio by using scalp distribution information.

The scalp distribution we filter for may not necessarily be that of the target component (in our case, P300). In fact, filtering for a different scalp distribution might produce even better results (i.e., improvement of the signal-to-noise ratio) than filtering for the target component. This may particularly be the case when some patterns of scalp distribution are more likely than others to be represented in the noise component. This condition is generally true for both systematic and background noise. In many cases noise activity at the central and parietal electrodes are strongly correlated. For this reason, a larger reduction of noise and

Off-line Frequency Filtering

The study included a comparison between five off-line low-pass frequency filtering procedures. All of them were based on a moving average filter (Ruchkin & Glaser, 1979). The procedure differed in the number of consecutive timepoints used for the smoothing (length), and in the number of iterations of the procedure adopted. In fact, two length levels (7 and 13 points, roughly equivalent to a 6.29 and a 3.14 Hz half cut-off filter respectively), and two iteration levels (1 and 2 iterations) were used. (Note that moving average filters cannot be perfectly described in terms of "half cut-off", because their frequency function is quite complex, as shown by Ruchkin & Glaser, 1979.) As a control, we used a condition where no off-line frequency filter was applied.

We want to emphasize that the comparison between filtering procedures described above does not exhaust all of the off-line frequency filters available to the investigator. We intend only to evaluate the effects of several frequency filters on latency estimations, and to determine whether, and to what extent, the general practice of smoothing waveforms improves the component latency estimation.

Spatial Filtering

Two spatial filtering procedures were compared: channel selection and Vector filtering.

Channel selection. This procedure consists of the selection of one electrode for further analysis. Given that our P300 component was maximum at Pz, we chose this electrode for the analysis.

Vector filtering. This procedure is based on the representation of

to that of P300. However, even in the worst case (large Slow Wave overlap) estimates obtained on Vector filtered data are no worse than those obtained on Pz waveforms.

For reasons of space, we cannot present here the results obtained with all the other combinations of component overlap, signal-to-noise ratio, frequency filtering, spatial filtering, and signal detection algorithm. It is sufficient to say that these results confirm the observations we have presented here.

In summary, the following conclusion can be made:

1. The error of latency estimation decreases exponentially as the signal-to-noise ratio increases.
2. Cross-correlation provides a more accurate estimate than peak picking.
3. Woody filter with 2 or 3 iteration is comparable to cross-correlation.
4. Frequency filters improve markedly the accuracy of estimates obtained with peak picking, and to a lesser, cross-correlation.
5. Vector filter yields estimates more accurate than channel selection (Pz).
6. Overlapping components impair the accuracy of estimates of channel selection, while the latency estimates of Vector filtered data are impaired only if the scalp distribution is similar to that of P300 (e.g. Slow Wave).
7. The accuracy improvement obtained with cross-correlation, Vector filter, and increase in signal-to-noise ratio appear to be independent (additive). The improvement in accuracy obtained

with Vector filter and frequency filtering are also independent.

8. The accuracy improvement obtained with cross-correlation and frequency filtering are not additive, that is, the combined use of both these procedures does not produce much better results than isolated use of either of them.
9. The scalp distributions of overlapping components differentially affect various spatial filtering procedures.
10. Latency adjustment procedures may "create" artifactual components. This is especially apparent at small signal-to-noise ratios. However, this is phenomenon is less evident when Vector filtered, rather than Pz, data are considered.

Additional analyses

Some of the findings listed above may be related to the particular conditions we used in our study. The procedures we adopted were mostly arbitrary (although we did attempt to simulate veridical conditions), and variations of some of the parameters may have a crucial impact on the accuracy of latency estimates. Thus, we ran three additional analyses to investigate the generalizibility of some of our findings. These three analyses explored the effects of standardizing background EEG noise, of variations in P300 wavelength, and of variations in the parameters used for Vector filter, on the accuracy of latency estimates.

Effect of standardizing background EEG noise. Our simulation of background EEG noise included the standardization, across trials and separately for each timepoint, of single trial deviations. The purpose was to obtain comparable variance across the whole epoch. However, this procedure might alter the veridicality of our simulation procedure. We

analyzed the frequency power spectra of the deviation waveforms before and after standardization. The comparison of the two average power spectra shown in Figure 2 did not reveal any particular alteration of the frequency characteristics of the deviation waveforms after standardization. To investigate further the effect of the standardization procedure, we ran part of the basic design of the study on non-standardized waveforms. The replication was exact, apart from the absence of frequency filtering. However, all the other manipulations were replicated. Note also that the signal-to-noise ratio for non-standardized waveforms could not be exactly determined. However, P300 amplitude was manipulated as in the basic design, and the level of noise in the P300 region was roughly comparable to that of the basic design. Thus, the same scale was adopted for the "signal-to-noise" ratio manipulation.

Some of the results obtained with non-standardized noise are presented in Figure 8.

Insert Figure 8 About Here

A comparison of the accuracy of latency estimation with standardized (cf. Figure 7, upper left panel) and non-standardized background noise (Figure 8) indicates that the standardization procedure did not significantly alter the results. All the findings were replicated. Thus, we may safely conclude that the standardization procedure did not impair the veridicality of the simulation.

Effect of P300 wavelength. The results from the basic design indicates that cross-correlation provides more accurate latency estimates than peak-picking, and that no further advantage is obtained by iterating the

Woody filter. However, the relatively high accuracy obtained with cross-correlation may be due to the fact that the template we used for this procedure exactly mirrors the target component (P300). This may also explain why adaptive templates, such as those used in the second and third iteration of Woody filter, do not produce any improvement. However, this situation may not be veridical. In reality, we may not know exactly the P300 waveshape, but only have some approximate estimate of it. Cross-correlation may be very sensitive to small differences between the template and the waveshape of the real component. On the other hand, it might be that the template which best discriminates between the target component and other sources of brain electrical activity does not mirror exactly the waveshape of the target component, but has some additional features which reduce its affinity with noise. Thus, we studied the accuracy of latency estimates obtained in conditions in which the template used for cross-correlation does not exactly correspond to the target component. We obtained this dissociation by varying the duration of P300.

To study the effect of P300 duration, we varied systematically P300 wavelength between 200 and 800 msec, with increments of 100 msec. However, we did not vary the wavelength of the template used for the cross-correlation procedure. P300 amplitude was fixed at 250 units (corresponding to a signal-to-noise ratio of 2.5). Background EEG noise was simulated through standardized waveforms, but no overlapping components were added. No frequency filtering was applied.

MSE (and number of trials required to obtain a standard error of 3 msec) as a function of P300 duration for two spatial filtering procedures (Pz selection and Vector filter) and four signal detection algorithm (peak picking, cross-correlation, Woody filter with two iterations, and Woody

filter with three iterations) are shown in figure 9.

Insert Figure 9 About Here

An inspection of this figure reveals several noteworthy findings. First, the accuracy of latency estimation depends on the duration of P300. The sharper the P300, the more accurate the estimate. This is particularly true for the peak picking procedure (especially if used in conjunction with Vector filter) and Woody filter. For cross-correlation, the most accurate estimation is obtained when the duration of the P300 is slightly shorter (400 msec) than that of the template. When the duration of the component is shorter than that of the template, peak picking and Woody filter produce estimates equal to or more accurate than cross-correlation. However, when the duration of the component is longer than the duration of the template, cross-correlation yields better estimates. These results suggest that peak picking and Woody filter produce accurate estimates in cases of sharp components. For peak picking, this is not surprising. For Woody filter, it may be that this procedure produces sharper templates at each iteration. Thus, iterating with Woody filter may be advantageous when the original template has a longer wavelength than the target component, but disadvantageous when the original template has a shorter wavelength than the target component. It is interesting to note that, in cases of latency "jitter", the averages tend to be "smooth", and the components "widened."

<u>Effect of variation of Vector filter parameters</u>. The parameters for Vector filter used in the basic design were chosen to discriminate between the target component (P300) and various sources of noise. We presented above the rationale for our choice. However, other parameters could have

been chosen, and some of them might have produced better results than those selected. To evaluate the consequences of the choice of Vector filter parameters, we varied them parametrically, and studied their relative impact on the accuracy of latency estimation.

For this study, we used a P300 amplitude of 250 units (signal-to-noise ratio). Background EEG noise was simulated with standardized waveforms, no overlapping component was added, and no frequency filtering was introduced. The parameters of Vector filter (polarity and orientation angles) were systematically varied by increments of 30 degrees. In particular, three levels of polarity (0, +30, +60 degrees) and twelve levels of orientation (-180, -150, -120, -90, -60, -30, 0, +30, +60, +90, +120, and +150 degrees) were used. The relative patterns of scalp distribution corresponding to some of these orientation angles, and the ratio between mean and standard deviation of the electrodes, corresponding to the polarity values, are shown in Figure 10. The effect of varying the parameters of Vector filter was evaluated by comparing it with the results obtained with channel selection (Pz). Thus, each of the 36 parameter combinations was classified as yielding estimates "clearly superior" (MSE more than 30% lower) than Pz selection, "superior" (MSE 0 to 30% lower) than Pz selection, "inferior" (MSE 0 to 30% higher) than Pz selection, and "clearly inferior" (MSE more than 30% higher) than Pz selection. These values are shown in Figure 10. This figure illustrates "regions" for which combinations of Vector filter parameters yield results clearly superior, superior, inferior, and clearly inferior to channel selection (Pz).

Insert Figure 10 About Here

Figure 10 shows that a specific region (combination of Vector filter parameters) yields the best results. This region does not correspond to the scalp distribution of P300. In fact, in this region the central electrode is more negative (or less positive) than the frontal electrode. The parameters of Vector filter we used for the basic design are within this region, although not at the center. Thus, the parameters we chose were "good". The MSE was lower at the center of the region than at the point corresponding to the parameters we used for the basic design. Thus, the parameters we chose on the basis of a previous empirical study were not the "best" possible for the simulated data.

Note that several combinations of Vector filter yield very low accuracy (regions where Vector filter is clearly inferior to Pz). This is not surprising. In fact, these combinations correspond to scalp distributions which do not enhance the discrimination between signal and noise. Rather, they enhance the noise or reduce the signal.

CONCLUSIONS AND GUIDELINES

The results obtained in this study indicate that the accuracy of latency estimation is affected by several variables, including the signal-to-noise ratio, characteristics of the signal and of the noise, the use of preparatory (filtering) procedures, and the choice of the signal detection algorithms.

The signal-to-noise ratio appears to be the main factor. In general, the error of estimation decreases exponentially with increases in the signal-to-noise ratio. Thus, any methodology which enhances the signal-to-noise ratio is strongly advocated. However, the effect of the signal-to-noise ratio does not appear to interact with other effects. In fact, procedures which yield the most accurate estimates at a high levels of signal-to-noise ratio, tend to produce the most accurate estimates at low levels of signal-to-noise ratio. Thus, the choice of the latency estimation procedure should not depend on the level of the signal-to-noise ratio at which the investigator is operating. While knowledge of the signal-to-noise ratio may be critical for estimating the <u>power</u> of the procedure, it is irrelevant for the <u>choice</u> of the algorithm for latency estimation.

The use of spatial information for enhancing the detection of ERP components can be very useful, particularly when some information about the scalp distribution of the target component is available. In general, Vector filter produced more accurate latency estimates than channel selection at Pz. This effect was most evident when overlapping components were present and when these components had a scalp distribution which was very different from the P300. However, the choice of the parameters of Vector filter is also important. These parameters should be such that the discrimination between signal and noise is enhanced, and not merely mirror the spatial characteristics of the signal. Although a mathematical algorithm for the correct selection of the Vector filter parameters is still lacking, a careful evaluation of the covariances among electrodes attributable to the component and to noise may be very useful. A possible solution may be the derivation of discriminant functions from "estimates" of the matrices of covariances among electrodes for component (signal + noise) and noise. In

the case of P300, the best results have been obtained with parameters which dissociate the parietal and the central electrode, while still emphasizing the positive trend across electrodes.

Frequency filtering produces improvement in the accuracy of latency estimation. The advantage is particularly evident when peak picking is used as the signal detection procedure. The accuracy obtained with cross-correlation does not seem to be much improved by the use of low-pass frequency filters, possibly because this procedure, in contrast with peak picking, is already based on several timepoints. Our investigation showed that the heavier the filtering, the more accurate the signal detection. Of course, there should be a level at which further filtering produces impairment of the signal detection because the signal itself is erased. Such a level of filtering was not reached in our study. It should be noted that our findings relate only to moving averages, since we did not compare the effect of filters with different band-pass characteristics, or the effect of high-pass filters, because of time and computational cost.

The choice of the signal detection algorithm may also affect the accuracy of our estimations. Cross-correlation produces better results than peak picking, at least when the wavelength of the template is comparable to, or shorter than, the wavelength of the signal. The difference between the two procedures may also be reduced by the use of appropriate frequency filters. Two or three iterations of Woody filter do not yield significant improvement over cross-correlation in cases in which the template for cross-correlation has a wavelength comparable to, or shorter than of the signal. However, the Woody filter iterations produced a marked improvement in accuracy in those cases in which the wavelength of the template was much longer (2 times or more) than that of the signal. Thus, cross-correlation

appears the best choice when the wavelength of the target component is known (at least approximately). When no information is available, Woody filter should be used. The use of peak picking should be restricted to the detection of sharp (wavelength equal or less than 300 msec) components, and, even in these cases, its use may be justified mainly on the basis of its simplicity and low computational load.

As a general commentary, the results of these studies emphasize that the characteristics of both signal and noise must be considered for the choice of procedures for the estimation of the latency of ERP components. The interaction between signal and noise characteristics was particularly evident for the choice of spatial filtering procedures. However, we believe that this is merely an instance of a general principle, and that the ERP signal detection algorithms should be based on those characteristics of the signal which allow its discrimination from the noise in which it is embedded.

Summary

We compared the accuracy of P300 latency estimation obtained with different procedures under different signal and noise conditions. The procedures included preparatory and signal detection techniques. Preparatory techniques were divided in frequency filters and spatial filters. In the latter category, we considered channel selection and Vector filter (Gratton, Coles, and Donchin, 1983). Signal detection techniques included peak picking, cross-correlation, and Woody filter with two or three iterations. P300 was simulated with a cosinusoidal wave (500 msec wavelength). Different signal-to-noise ratios were simulated by multiplying

the signal by a scaling factor. Two kinds of noise were added: event-related noise (overlapping components), and non-event related noise (background EEG). The first was simulated by adding to the signal a set of four cosinusoidal waves (one complete cycle). Background EEG noise was simulated by obtaining single trial records from an odddball experiment "free" of time locked activity. The signal and noise conditions were systematically varied in a factorial design yielding 20 noise x 11 signal conditions. In addition, 5 frequency filter conditions, 2 scalp distribution filter conditions, and 4 signal detection algorithms were also applied factorially. The accuracy of the different procedures was estimated using root mean square error estimates.

Several effects were noteworthy. Accuracy increased exponentially as a function of the signal-to-noise ratio. Cross-correlation was advantageous in comparison with peak-picking. The results with Woody filter parallels those obtained with cross-correlation. Vector filter was advantageous in comparison with channel selection. The use of frequency filtering reduces the advantage of cross-correlation, but not the effect of increasing the signal-to-noise ratio, or the advantage of Vector filter. The conditions of large component overlap impaired the accuracy of the estimates obtained with channel selection. They impaired the accuracy of the estimates obtained with Vector filter only when the overlapping component had a scalp distribution similar to that of the signal component. The effects of varying noise characteristics, P300 duration, and parameters of Vector filter were also investigated.

These results point to the advantage of using all the available information in the estimation of ERP components latency. This information may concern both the time course and the scalp distribution of the signal

component. Characteristics of the noise are also relevant to the choice of the appropriate procedure. Some guidelines for this choice were also provided.

Table 1. Parameters of overlapping components

Component	Amplitude (units)	Latency (msec)	Duration (msec)	Scalp Distribution Weights Fz	Cz	Pz
N100	100	100	100	-.6	-1.8	-.6
P200	150	250	150	1.0	1.0	.5
N200	50/100	300/400	250	-1.2	-.8	-.4
SW	100/200	800/1080	800	-.4	.0	.4
P300*	0 to 500 (by 50's)	550	500	.4	.8	1.2

* P300 is included for comparison

Table 2. List of experimental conditions

(A) Trials per condition (100)

(B) Signal detection algorithms (4)

peak-picking
cross-correlation
Woody with two iterations
Woody with three iterations

(C) Scalp distribution filtering techniques (2)

channel selection (Pz)
Vector filtering

(D) Target component amplitude levels (11)

Signal-to-noise ratio varying from 0 to 5 by .5 jumps

(E) Amplitude levels of overlapping components (5)

N200: 50, 100
SW: 100, 200
no overlap

(F) Latency levels of overlapping components (4)

N200: 300, 400
SW: 800,1080

(G) Frequency filter conditions (5)

6.29 Hz, 1 iteration
6.29 Hz, 2 iterations
3.14 Hz, 1 iteration
3.14 Hz, 2 iterations
no filter

(H) Dependent Variables (2)

Mean square error of estimation
Number of trials required to obtain 3 msec error

Fig. 6

Signal-to-noise ratio
(Amplitude of P300)

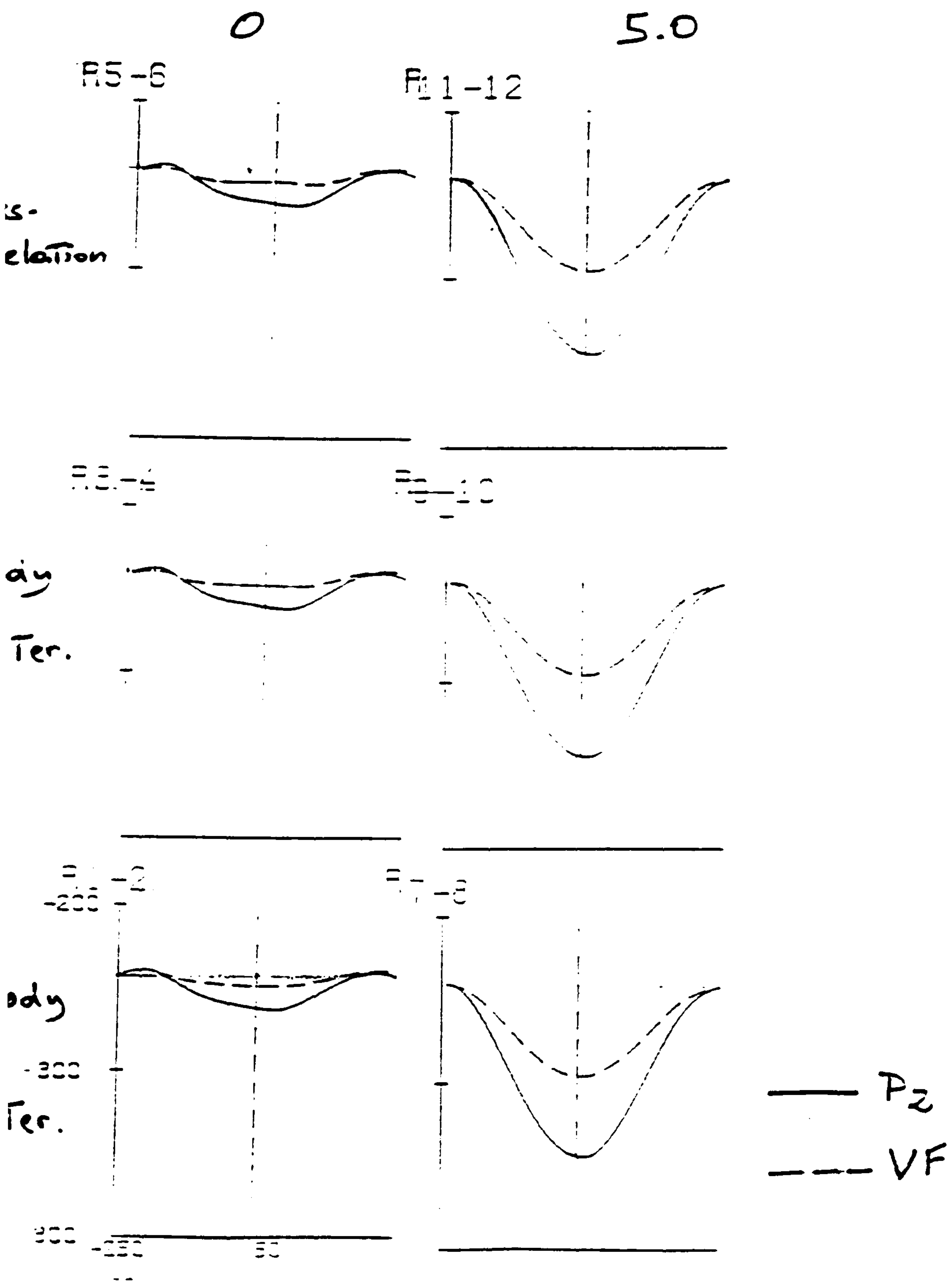
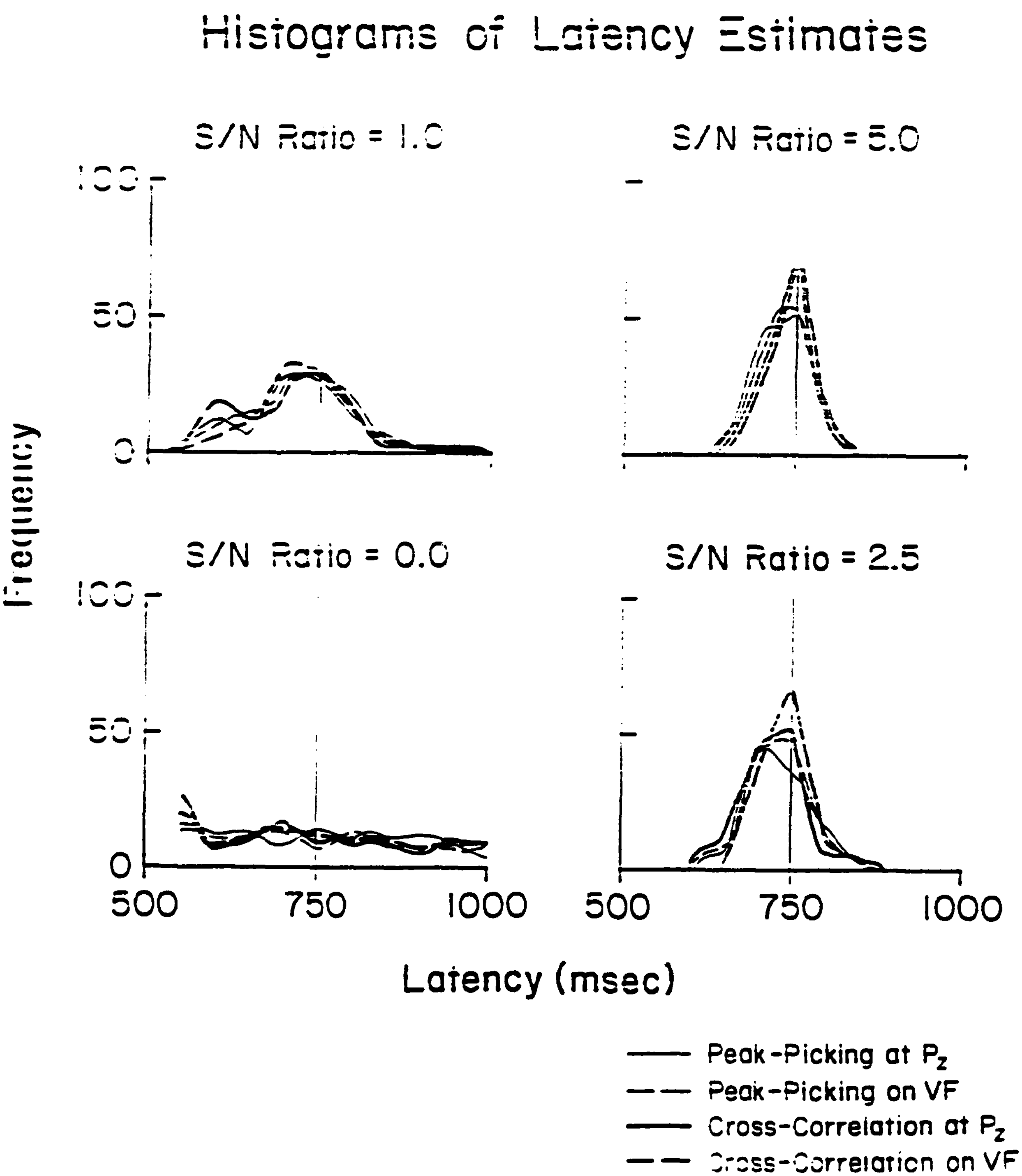
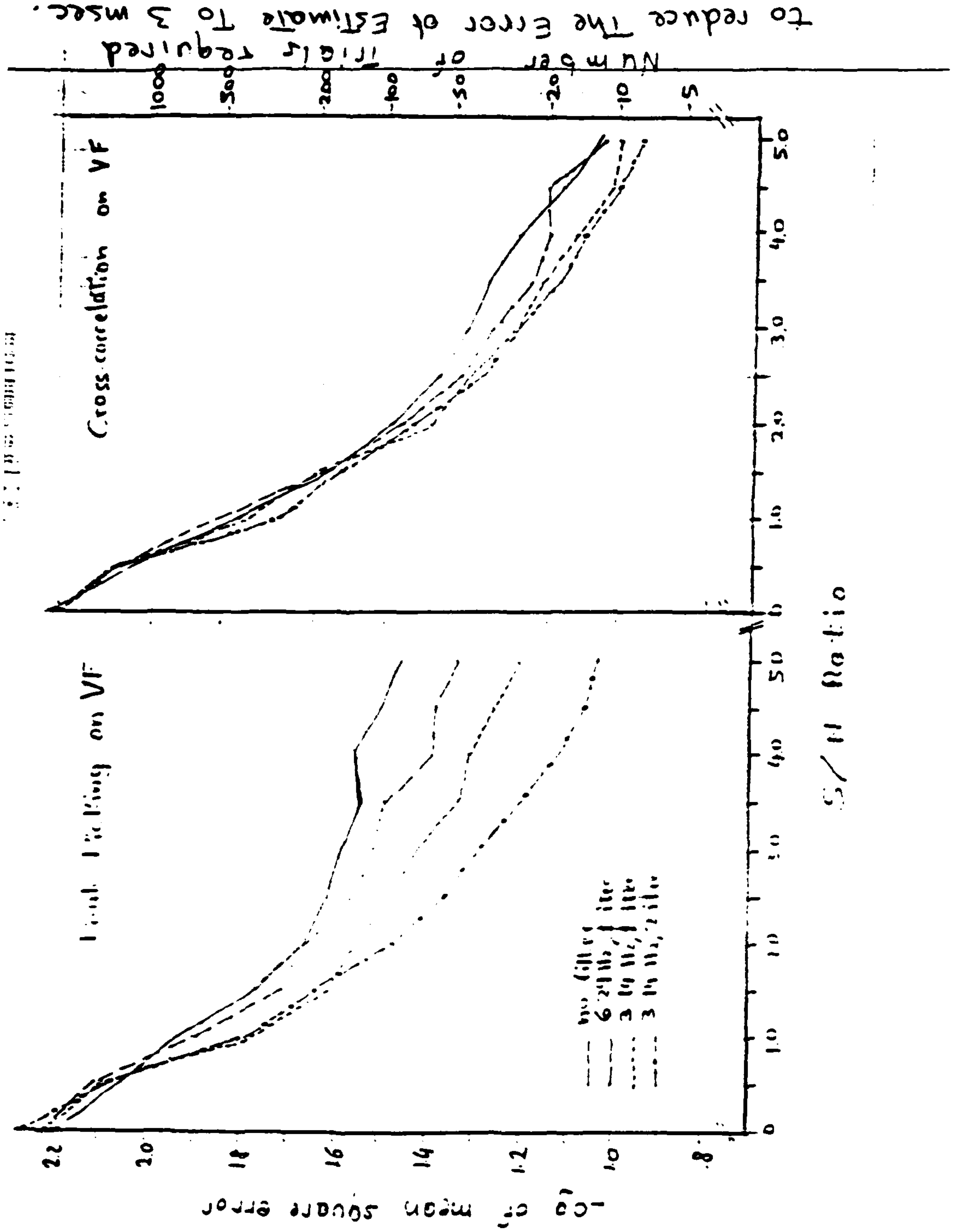


Fig. 5



Histograms of Latency Estimates
S/N Ratio = 1.0
S/N Ratio = 5.0
S/N Ratio = 0.0
S/N Ratio = 2.5
Frequency
100
50
0
500
750
1000
Latency (msec)
Peak-Picking at P_z
Peak-Picking on VF
Cross-Correlation at P_z
Cross-Correlation on VF

Fig. 4b



Number of Trials required to reduce The Error of Estimate To 3 msec.
1000
500
200
100
50
20
10
5
Cross-correlation on VF
S/N Ratio
-Log of mean square error
2.2
2.0
1.8
1.6
1.4
1.2
1.0
.8

Effect of Filtering on Latency Estimation

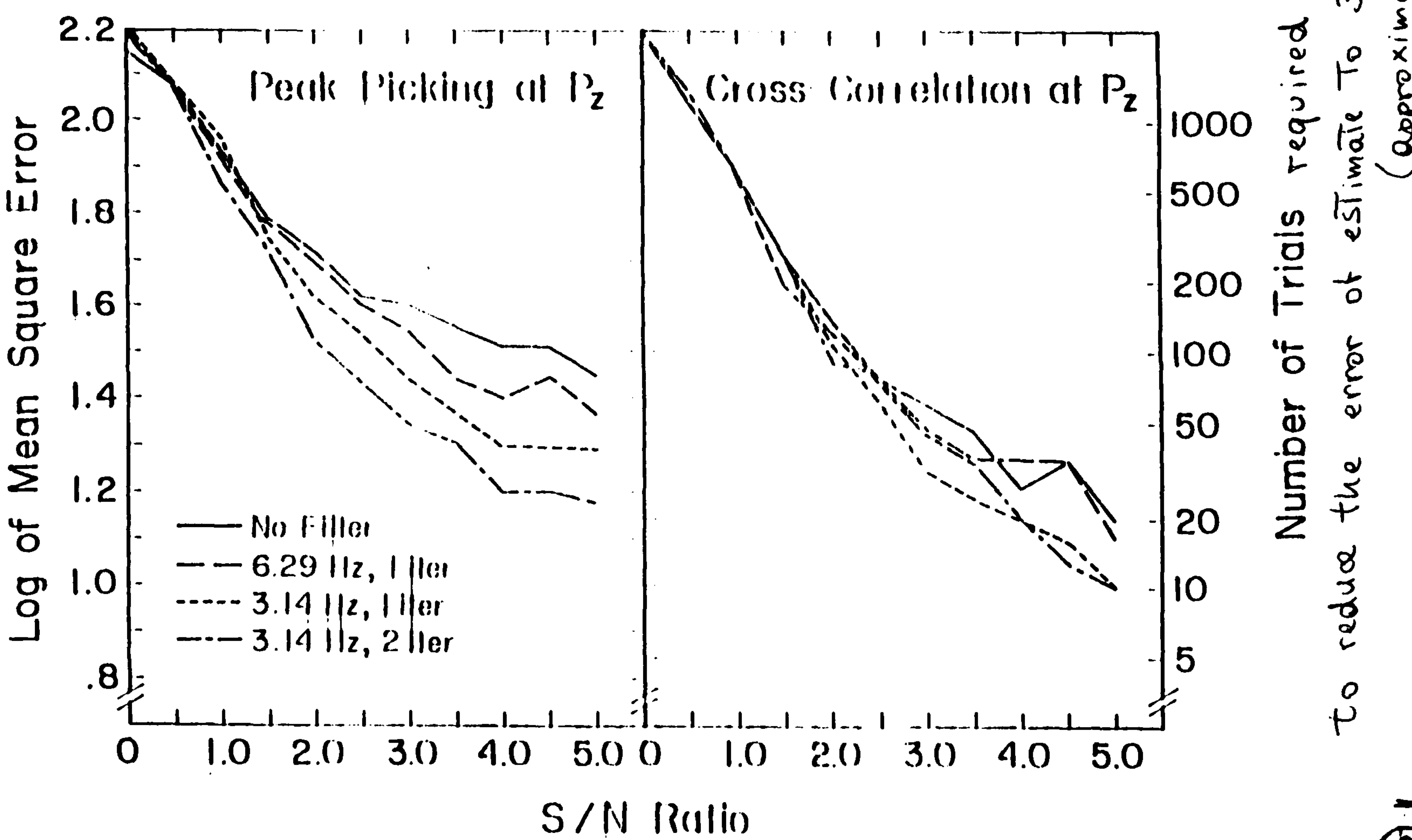
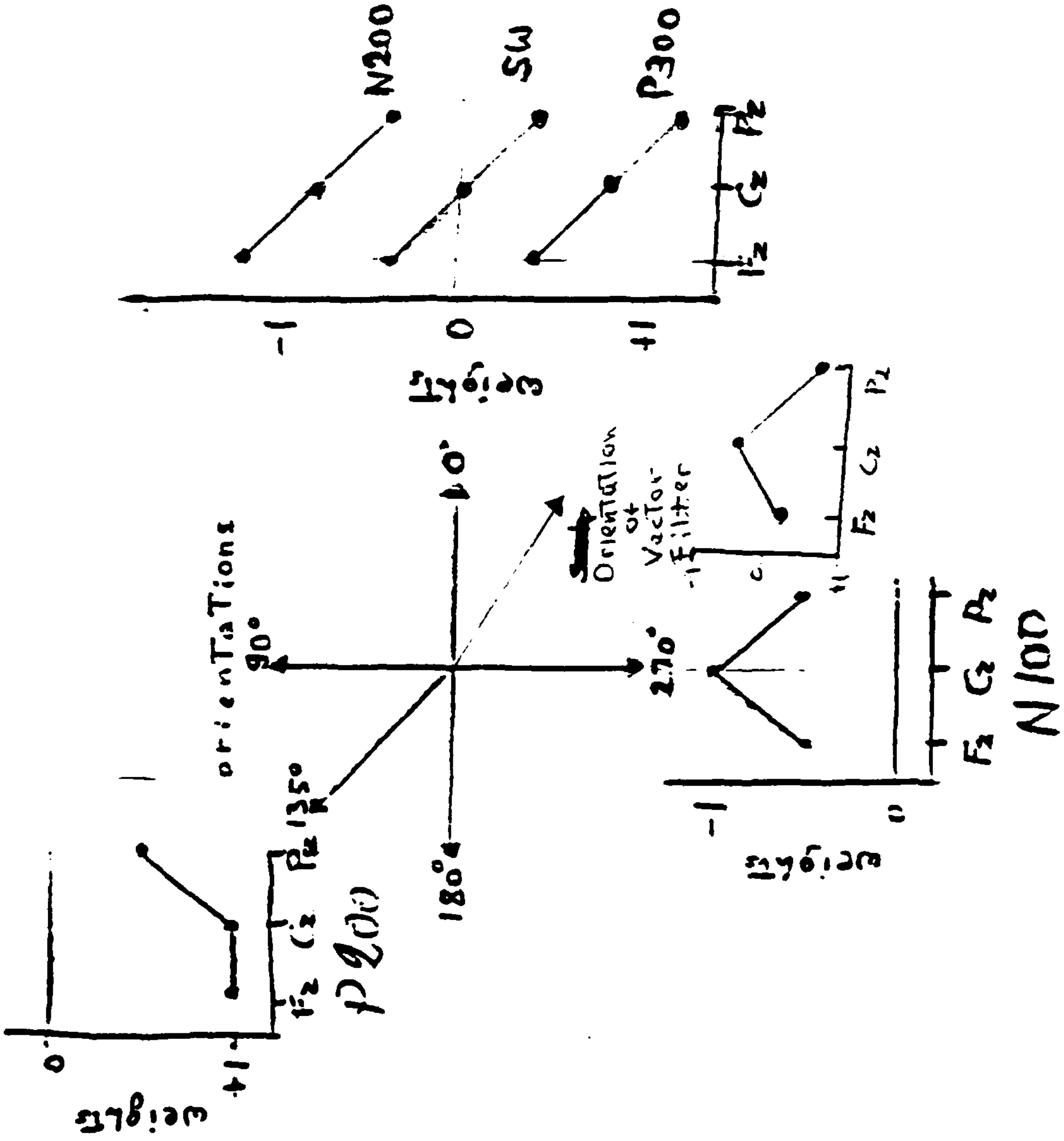


Fig. 4a

Fig. 3



N200
SW
P300
Fz
Cz
Pz
-1
0
+1
weight
orientaTions
90°
0°
180°
270°
135°
Orientation of Vector Filter
N100
P2(3)

AVERAGE POWER SPECTRA DENSITY

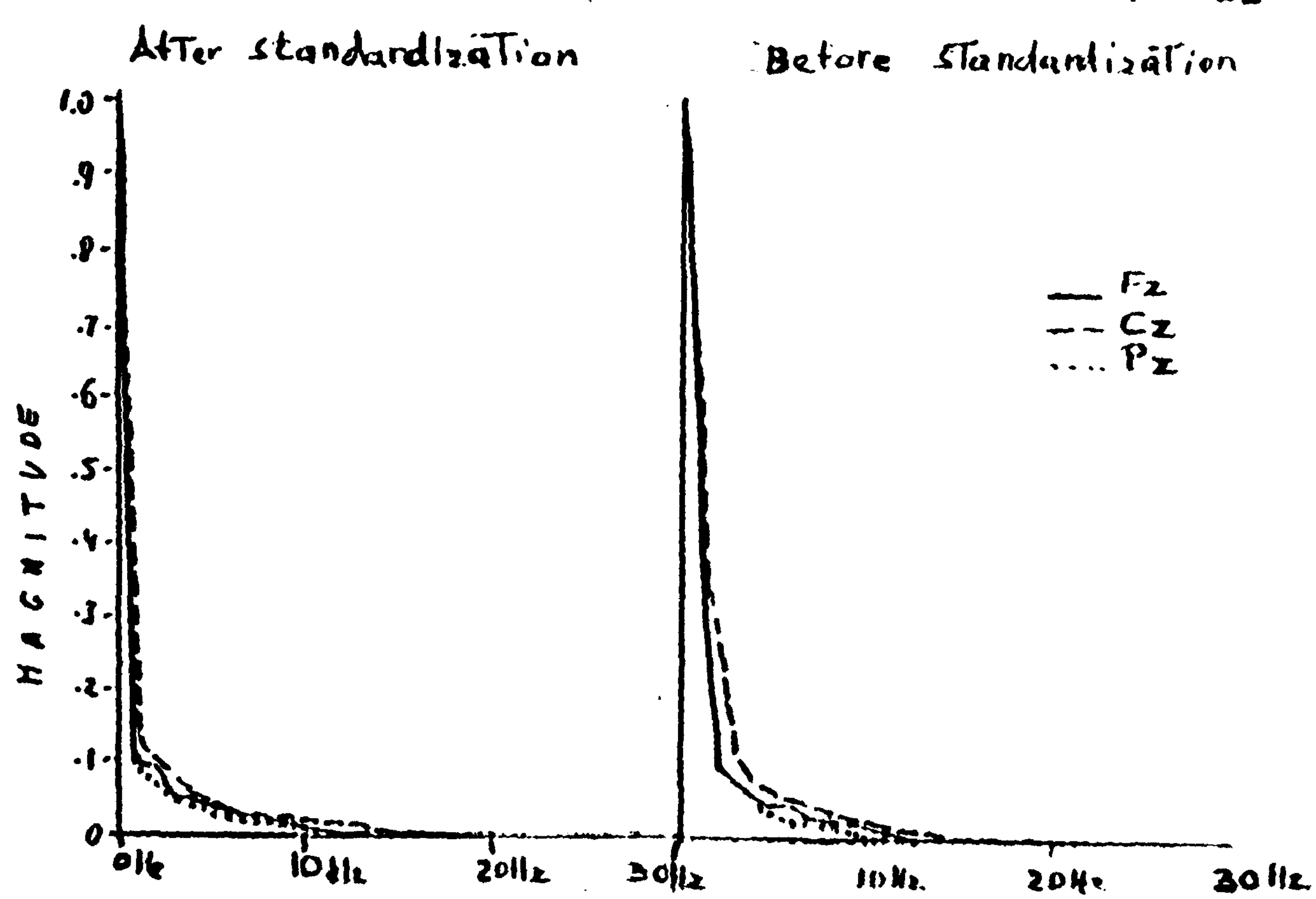


Fig. 2b

Fig. 2a

Fig. 1

Examples of Simulated Waveforms

(P300 Amplitude = 250)

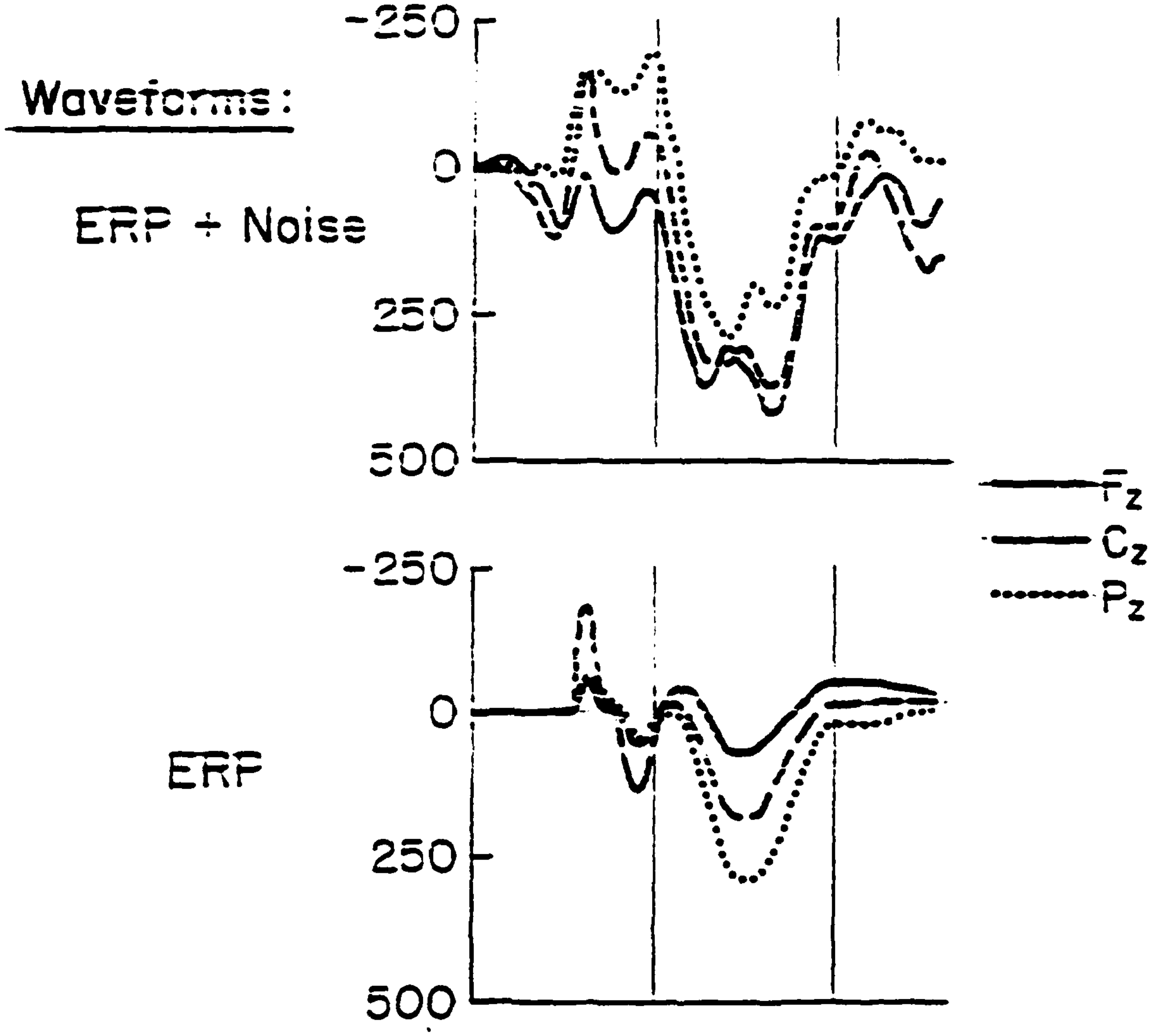


Figure 7. Log mean square error (and number of trials required to obtain a standard error of 3 msec) as a function of signal-to-noise ratio for four different component overlap conditions. Two detection algorithms and two spatial filtering techniques are shown for each condition.

Figure 8. Log mean square error (and number of trials required to obtain a standard error of 3 msec) as a function of signal-to-noise ratio for the non-standardized noise condition. Two detection algorithms and two spatial filtering techniques are shown.

Figure 9. Effect of P300 duration on the accuracy of latency estimation with peak-picking, cross-correlation and Woody filter with two and three iterations.

Figure 10. Effect of the manipulation of Vector Filter parameters on the accuracy of latency estimation as compared to the use of a single electrode (Pz).

Figure Legends

Figure 1. Examples of simulated waveforms from a single trial. Waveforms with simulated EEG noise are shown at the top, waveforms without noise are displayed at the bottom of the figure.

Figure 2. Average power spectra of the simulated background noise. (a) Non-standardizes background noise: (b) Standardized background noise.

Figure 3. Orientation angles corresponding to the scalp distribution of the components used in the study.

Figure 4. Log mean square error (and number of trials required to obtain a standard error of 3 msec) as a function of signal-to-noise ratio, for different frequency filters, signal detection algorithms, and spatial filtering conditions.

Figure 5. Histogram of latency estimates for four different signal-to-noise ratios from two detection algorithms and two spatial filtering techniques.

Figure 6. Latency adjusted average waveforms over 1[illegible] trials. P300 peak latency was computed with cross-correlation (lower panels), Woody 2-iterations (middle panels), Woody 3-iterations (upper panels). The left column refers to waveforms obtained with a signal-to-noise ratio of 0 (no P300 was present), the right column refers to waveforms obtained with a signal-to-noise ratio of 5. The solid lines indicate Pz waveforms, the dashed lines indicate Vector filtered waveforms.

Ungar, P., & Basar, E. Comparison of Wiener filtering and selective averaging of evoked potentials. Electroencephal. clin. Neurophysiol., 1976, 40: 516-520.

Wastell, D.G. Statistical detection of individual evoked responses: an evaluation of Woody's adaptive filter. Electroencephal. clin. Neurophysiol., 1977, 42: 835-839.

Wastell, D.G. When Wiener filtering is less than optimal: an illustrative application to the brainstem evoked potential. Electroencephal. clin. Neurophysiol., 1981, 51: 678-682.

Wiener, N. Extrapolation, interpolation and smoothing of stationary time series. John Wiley, New York, 1949.

Woody, C.D. Characterization of an adaptive filter for the analysis of variable latency neuroelectrical signals. Medical and Biol. Eng., 1967, 5: 539-553.

Nahvi, M.J., Woody, C.D., Ungar, R., & Sharafat, A.R. Detection of neuroelectrical signals from multiple data channels by linear filter method. Electroencephal. clin. Neurophysiol., 1975, 38: 191-198.

Palmer, C.W., Derbyshire, A.J., & Lee, A.W. A method for analysing individual cortical responses to auditory stimuli. Electroencephal. clin. Neurophysiol., 1966, 204-206.

Pfefferbaum, A. A simulation study of latency adjustment procedures. Paper presented at III Workshop on ERP Methodology, Annual Meeting of the Society of Psychophysiological Research, Asilomar, California, 1983.

Pritchard, W.S. Psychophysiology of P300. Psychological Bulletin, 1981, 89: 506-540.

Ruchkin, D.S., & Glaser, E.M. Some simple digital filters for examination of CNV and P300 on single trial basis. In D. Otto (Ed.) Multidisciplinary perspectives in event-related brain potentials. Government Printing Office, 1979: 579-581.

Squires, K.C., Wickens, C., Squires, N.K., & Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 1976, 193: 1142-1146.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. P300: methodological and theoretical issues. In P.K. Ackles, J.R. Jennings, and M.G.H. Coles (Eds.), Advances in psychophysiology. Vol. II. JAI Press, Greenwich, CT, in press.

Gratton, G., Coles, M.G.H., & Donchin, E. A new method for off-line removal of ocular artifact. Electroencephal. and Clin. Neurophysiol., 1983a, 55: 468-484.

Gratton, G., Coles, M.G.H., & Donchin, E. Filtering for scalp distribution: a new approach (Vector filter) (Abstract). Psychophysiol., 1983b, 20: 443-444.

Gratton, G., Coles, M.G.H., & Donchin, E. Component identification with Vector analysis. Paper presented at the III International Conference on Cognitive Neuroscience, Bristol, England, 1984.

Gratton, G., Coles, M.G.H., & Donchin, E. Vector Analysis of event-related brain potentials. In preparation.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197: 792-795.

McCarthy, G., & Donchin, E. Brain potentials associated with structural and functional visual matching. Neuropsychologia, 1978, 18: 571-585.

Government Printing Office, 1979: 555-572

Donchin, E., & Herning, R.I. A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event related potentials. Electroencephal. and Clin. Neurophysiol., 1975, 38: 51-68.

Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: the endogenous components of the ERP. In E. Callaway, P. Tueting, and S.H. Koslow (Eds.) Event-related brain potentials in man. Academic Press, New York, 1978.

Duncan-Johnson, C.C. P300 latency: a new metric for information processing. Psychophysiol., 1981, 18: 207-215.

Duncan-Johnson, C.C., & Donchin, E. The time constant in P300 recording. Psychophysiol., 1979, 16: 53-55.

Duncan-Johnson, C.C., & Donchin, E. The P300 component of the event-related brain potential as an index of information processing. Biol. Psychol., 1982, 14: 1-52.

References

Coles, M.G.H., & Gratton, G. Psychophysiology and contemporary models of human information processing. In D. Papakostopoulos, and I. Martin, (Eds.), Clinical and Experimental Neuropsychophysiology. Croom Helm Ltd., Beckenham, England, in press.

Coles, M.G.H., Gratton, G., Kramer, A.F., & Miller, G.A. Principles of signal acquisition and analysis. In M.G.H. Coles, E. Donchin, and S.W. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. Guilford Press, New York, in press.

Derbyshire, A.J., Driessen, G.J., & Palmer, C.W. Technical advances in the analysis of single acoustically evoked potential. Electroencephal. clin. Neurophysiol., 1967, 22: 467-481.

Donchin, E. Surprise?...Surprise! Psychophysiol., 1981, 18: 493-513.

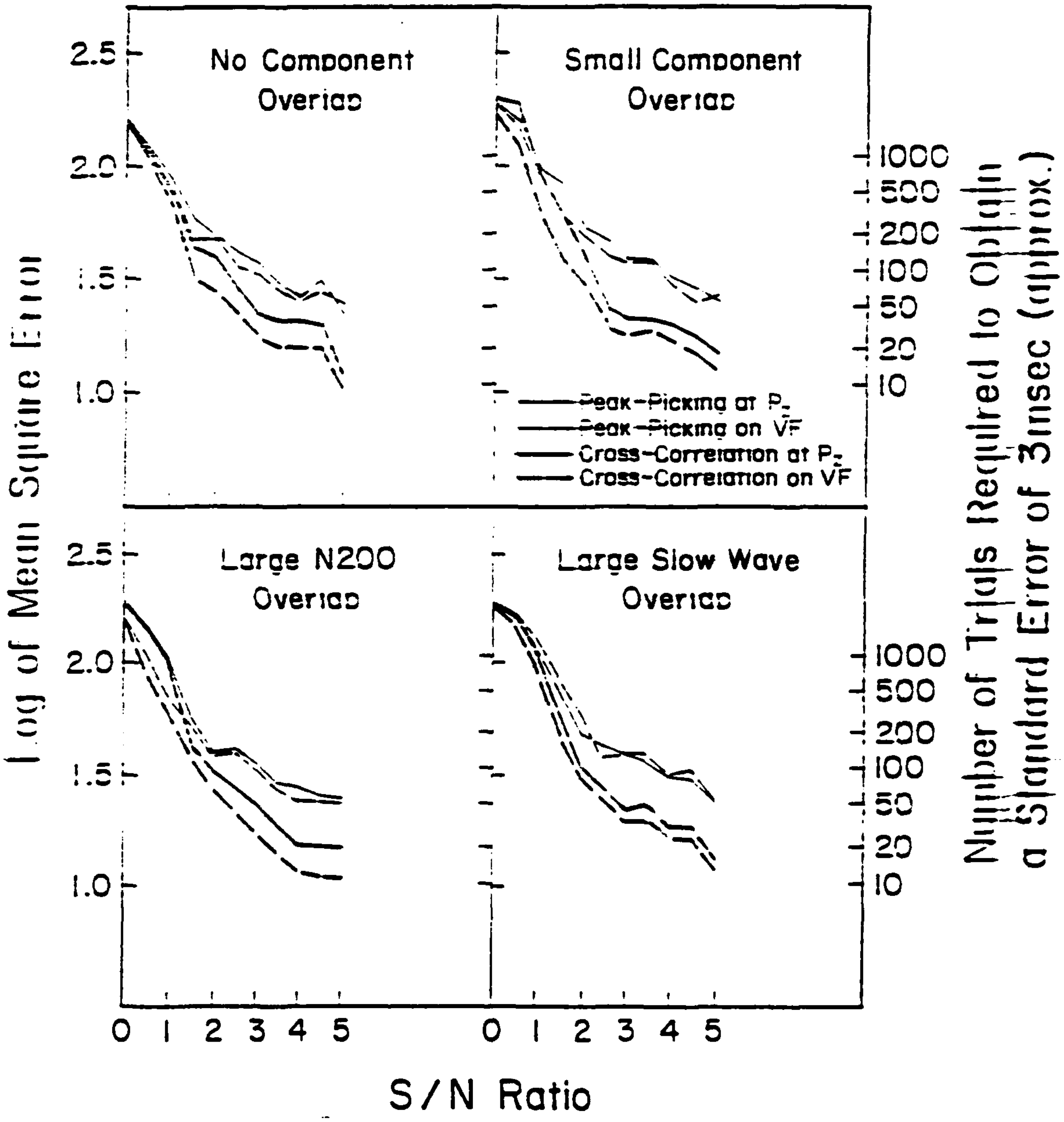
Donchin, E., Coles, M.G.H., & Gratton, G. Cognitive psychophysiology and preparatory processes: a case study. In S. Kornblum and J. Requin (Eds.), Preparatory state and processes. LEA, Hillsdale, NJ, 1984: 155-178.

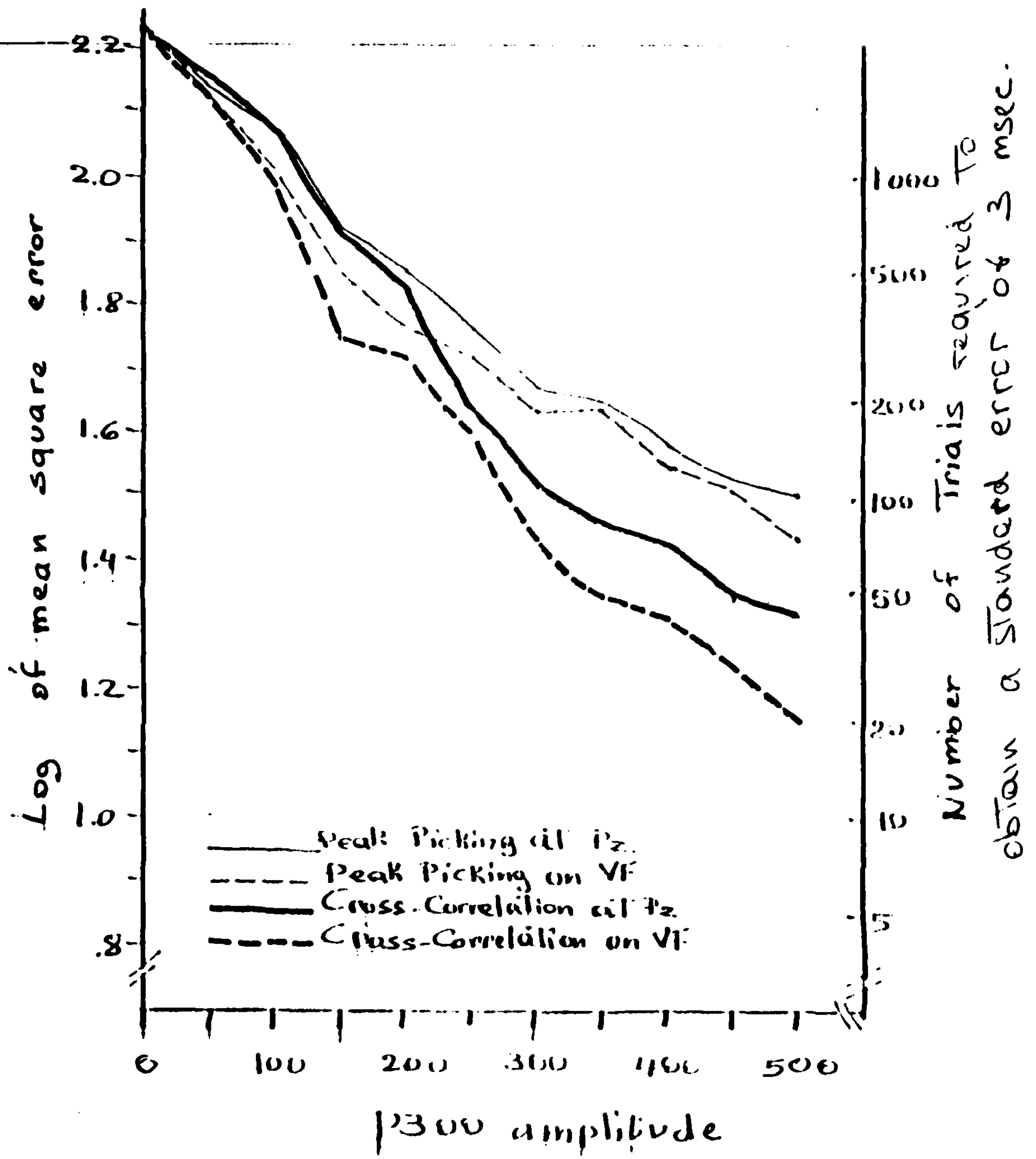
next page

Donchin, E., & Heffley, E.F. Multivariate analysis of event-related potential data: a tutorial review. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potentials.

Fig.7

Effect of Component Overlap on Latency Estimation



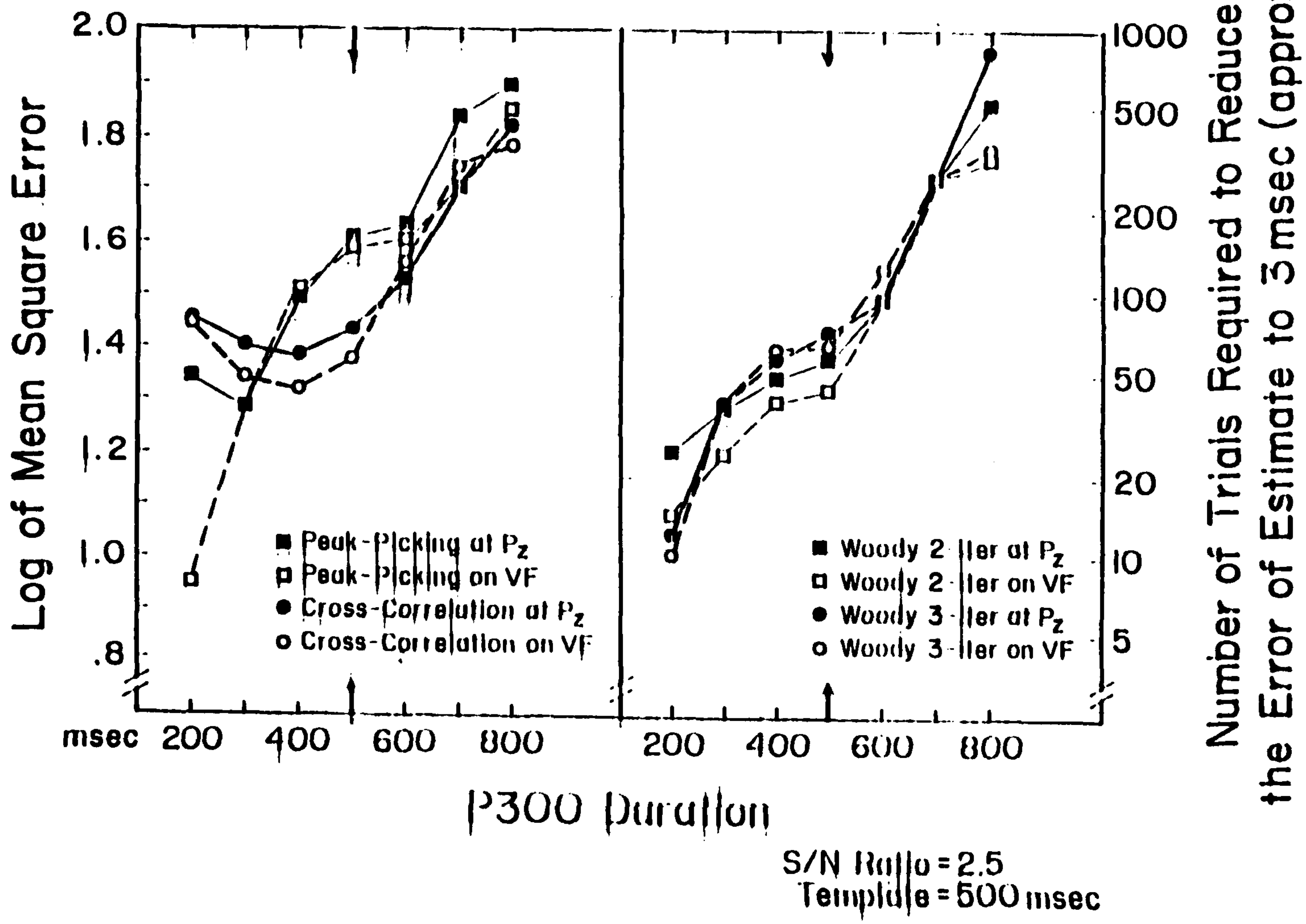


Log of mean square error
2.2
2.0
1.8
1.6
1.4
1.2
1.0
.8
Number of Trials required to obtain a standard error of 3 msec.
1000
500
200
100
50
20
10
5
Peak Picking at Pz
Peak Picking on VF
Cross-Correlation at Pz
Cross-Correlation on VF
0
100
200
300
400
500
P300 amplitude

Fig. 8

Fig. 9

Error of Latency Estimation as a Function of P300 Duration



Effect of Manipulation of the Parameter of Vector Filter

Black = "Much" better than P_z (more than 30%)
Gray = "Slightly" better than P_z (from 0 - 30%)
White = "Worse" than P_z

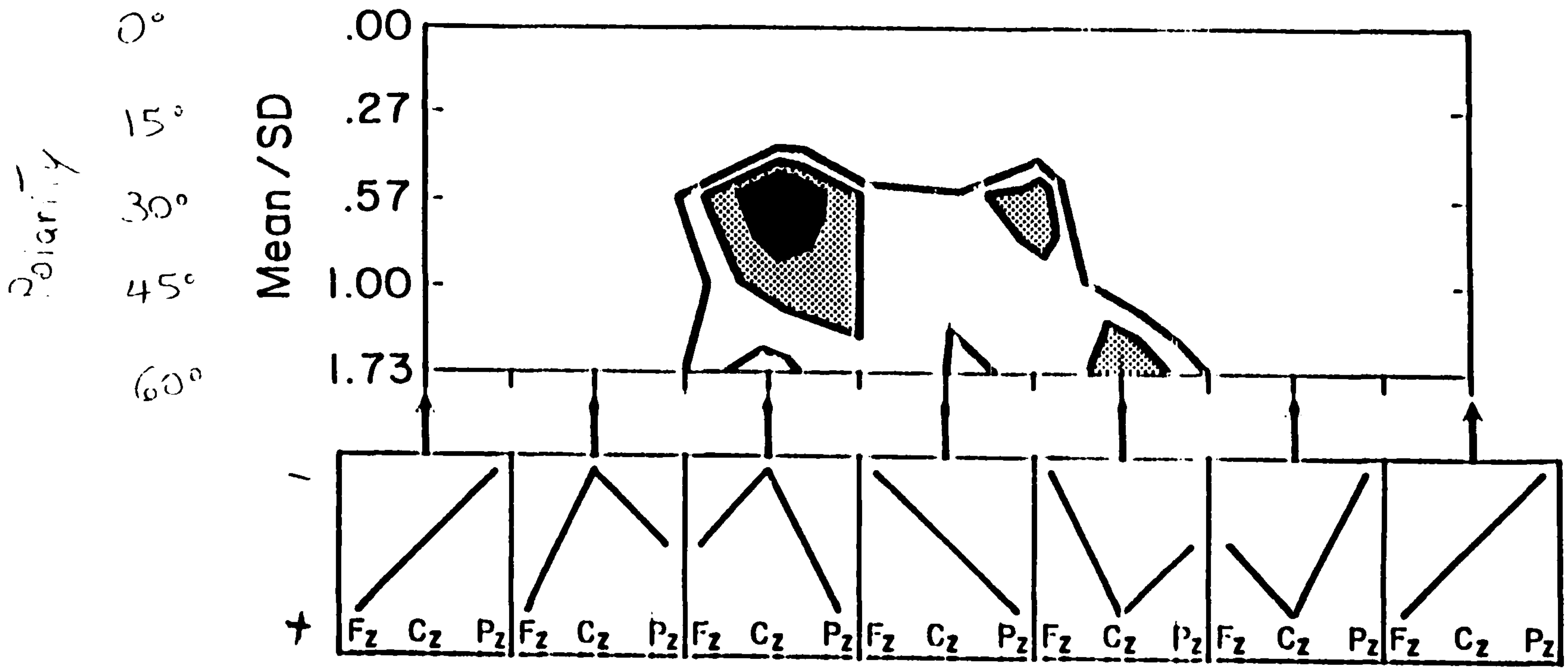


Fig. 10

Psychophysiology, in press.

The Effects of Practice and Task Structure on Components of the Event-Related Brain Potential

Arthur Kramer, Walter Schneider, Arthur Fisk & Emanuel Donchin

Department of Psychology
University of Illinois
Champaign, Illinois

Acknowledgments

This research was supported by the Air Force Office of Scientific Research under contract No. F49620-79-C-0233 with Dr. Alfred Fregly as technical monitor. We gratefully acknowledge the helpful comments of Gabriele Gratton, Mick Rugg and David Strayer.

Requests for reprints should be sent to Arthur F. Kramer, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel, Champaign, Illinois, 61820.

Running Title: ERPs, Task Structure and Practice

Abstract

The present study focused on the effects of, and the interactions between, practice and task structure on human performance. The development of automatic processing through consistent stimulus-response mapping (CM) was assessed by means of measures of reaction time and event-related brain potentials. The subjects performed a visual search task in which they responded by pressing a button whenever a probe matched a memory set item. The variables manipulated in the study included the number of memory set items (1 or 4), the task structure (CM or VM), and the probability of occurrence of a memory set item (.2 or .8). Set size had a significant effect on RT in both CM and VM conditions prior to practice and in the VM condition following extensive practice. P300 latency mirrored RT, suggesting that the development of automatic processing substantially reduced stimulus evaluation time. The commonly observed relation between probability and P300 amplitude, with larger P300s elicited by infrequent events, was found in the VM conditions but not in the CM condition after practice. This suggests an attenuation of memory updating during automatic processing. Two different negative components were affected by stimulus mismatch. These components appear to reflect different degrees of mismatch processing.

Descriptors: Event-related brain potentials (ERP), automatic and controlled processing, visual search, task structure, practice.

The Effects of Practice and Task Structure on Components of the Event-Related Brain Potential

Arthur Kramer, Walter Schneider, Arthur Fisk & Emanuel Donchin

The interaction of extensive practice and task structure have dramatic effects on subjects' performance. A two process theory proposed by Schneider and Shiffrin (1977) postulated that the consistency of stimulus-response mapping was responsible for the type of information processing performed by subjects. A subject will develop an automatic processing response if during a training period, stimuli are consistently mapped to responses. The automatic processing response is characterized as fast, inflexible, difficult to suppress once learned, and not limited by short-term memory capacity. Automatic processes require extensive training to develop but once acquired can be performed concurrently with other tasks with little if any performance deficit. The automatic process appears to be graded rather than all-or-none. Tasks with high levels of stimulus-response consistency result in greater improvements with practice than tasks with intermediate levels of consistency (Schneider & Fisk, 1982). Controlled processing occurs in situations in which subjects are unable to consistently map stimuli to responses. Controlled processing is characterized as slow, serial and capacity limited. Asymptotic controlled processing develops with little training and the processing is modifiable by the subject. The qualitative and quantative differences between automatic and controlled processing modes have been demonstrated in character (Shiffrin & Schneider, 1977), word, and category search tasks (Fisk & Schneider, 1983; Schneider & Fisk, 1984). In the present study the effects of these two modes of processing on the components of the Event-Related Brain Potential (ERP) will be examined.

The search task commonly employed in the study of automatic and

controlled processing requires subjects to memorize a set of items and later compare a set of visually presented items to members of the memory set (Sternberg, 1966, 1969). Subjects make one response (target present) if one of the visually presented items was from the memory set and another response (target absent) if the items do not match any of the members of the memory set. Memory set items are labeled targets; items not included in the memory set are referred to as distractors. Automatic processing develops in a consistent mapping condition (CM) in which targets are always selected from one category (e.g. letters A to M) and distractors are chosen from another category (e.g. letters N to Z). Thus, a target present response is always produced for one set of items and a target absent response for the other set. Controlled processing is employed in a varied mapping condition (VM) in which subjects are unable to consistently map stimuli to responses. In the VM condition both targets and distractors are chosen from the same category (e.g. letters A to Z). Targets and distractors exchange roles over trials in the VM condition.

At least two criteria have been used to define automatic processing. In the varied mapping condition of the visual search task, reaction time (RT) is a linear function of the size of the memory set, increasing by approximately 40 msec with the addition of each alphanumeric character. However, after extensive practice in the CM condition the memory set slope decreases until the addition of a memory set item fails to produce any increase in RT. This effect suggests that the fully developed automatic process is not limited by short term memory. The flat memory set slope is one criterion used to define "automaticity" of processing. Several recent studies have demonstrated that automatic processing tasks can be performed concurrently with controlled processing tasks without incurring any performance deficits (Schneider & Fisk, 1983, 1984). In one such study

subjects concurrently performed two visual search tasks (Schneider & Fisk, 1982). One of the tasks was performed with consistent stimulus-response mapping while in the other task subjects were unable to consistently map stimuli to responses (VM condition). As long as performance of the controlled processing task was emphasized, subjects were able to execute both tasks together without a performance deficit. The absence of a dual-task decrement suggests that the automatic task can be performed without consuming processing resources. This dual-task effect represents the second criterion for automaticity of processing.

The dependent variables employed in the study of automatic and controlled processing have included RT, percent correct and measures of signal detection theory, d' and beta. Although these measures provide a reliable representation of the accuracy and latency of a subjects response in the visual search task, they do not permit the investigators to easily decompose processes which take place between the encoding of the stimulus and the execution of the response. Several components of the ERP have been shown to be sensitive to a subset of the processes reflected by RT and are therefore useful in augmenting the chronometric information provided by more traditional measures of performance.

Event-Related Brain Potentials

The P300 component of the ERP is a positive going deflection in the ongoing EEG which occurs with a minimal latency of 300 msec following the presentation of a task relevant event. A task in which the P300 is readily elicited is often called the "oddball" paradigm. In a study by Duncan-Johnson and Donchin (1977) using this paradigm, subjects were instructed to covertly count one of two tones that were presented in a Bernoulli series of high and low tones. In different blocks of trials the relative probability of the two tones was manipulated. The amplitude of the

P300 increased monotonically as the probability of the stimulus decreased. This occurred regardless of which of the two tones was being counted.

The demonstration that P300 is elicited by unexpected, task relevant stimuli led Donchin and Isreal (1980) to suggest that, "the P300 reflects the activation of a processing mechanism which is engaged whenever it is necessary to update our mental model of the environment" (see also Donchin, McCarthy, Kutas & Ritter, 1983). The frequency with which the model is revised is based on the surprise value and task relevance of the stimuli. A recent study has strengthened the argument that P300 reflects the updating of working memory. Karis, Fabiani and Donchin (1984) presented subjects with lists of words they were to memorize. For subjects who used rote mnemonic strategies, the amplitude of the P300 elicited by words during their initial presentation predicted later recall. Larger P300s were elicited by words which were later recalled. In the present study, the effects of task structure and practice on the P300 probability effect will be examined. Once subjects have developed an automatic processing response the continual updating of working memory becomes unnecessary (Shiffrin & Schneider, 1977; Fisk & Schneider, 1984). On the basis of this argument it is predicted that the probability effect which is presumed to represent the updating of a mental template (Donchin, 1981) will be reduced in the CM condition following extensive practice.

The peak latency of the P300 component has been found to be influenced by stimulus evaluation processes while being relatively unaffected by the processes of response selection and execution (Maligero, Bashore, Coles & Donchin 1984; Squires, Donchin, Herning & McCarthy, 1977). McCarthy and Donchin (1981) orthogonally manipulated two independent variables in an additive factors design. One factor, stimulus discriminability, has been shown to affect an early encoding stage of processing while the second

factor, stimulus-response compatibility, influences the later stages of response selection and execution. The subjects' task was to decide which of two target stimuli, the words RIGHT or LEFT, were presented in a matrix of characters on a CRT. Stimulus discriminability was manipulated by varying the amount of noise in the matrix. Stimulus-response compatibility was manipulated by requiring subjects to respond to the target stimulus with the compatible or incompatible hand (i.e. compatible - respond to the word RIGHT with the right hand). RT increased when the target word was embedded in noise and when the response was incompatible with the stimulus. P300 latency was increased by the addition of the noise to the target matrix, but was not affected by the incompatibility between the stimulus and response. Thus, the results support the conclusion that P300 latency is affected by a subset of the processes which affect RT. Additional evidence of P300's sensitivity to stimulus evaluation processes has been obtained in varied mapping visual search paradigms. In these studies, both RT and P300 latency increase monotonically with increasing memory load (Adam & Collins, 1978; Brookhuis et al., 1983; Ford et al., 1979, 1980; Gomer, Spicuzza & O'Donnell, 1976). This selective sensitivity of the P300 to a subset of information processing activities has been found useful in augmenting the chronometric information provided by measures of RT (Duncan-Johnson, 1981; Duncan-Johnson & Donchin, 1982). In the present study, the latency of P300 is expected to reflect the development of automaticity in the consistent mapping conditions. It is predicted that the slope of P300 latency as a function of memory set size will decrease substantially in the practiced CM conditions indicating an automization of stimulus evaluation processes.

Another component of the ERP, the N200, has been found to be a reliable indicator of stimulus mismatch (Naatanen, Simpson & Loveless, 1982; Naatanen & Gaillard, 1983). The N200 appears to be an automatic response to stimulus

mismatch since it occurs regardless of the focus of subjects attention (Naatanen, Gaillard & Varey, 1981; Squires et al., 1977). N200's have been elicited by mismatches in visual, auditory and somatosensory modalities in a diverse set of experimental paradigms including selective attention tasks, omitted stimulus paradigms, lexical decision tasks and word matching paradigms (Ford, Pfefferbaum & Kopell, 1982; Ford, Roth & Kopell, 1976; Klinke, Fruhstorter & Finkenzeller, 1968; Kramer, Ross & Donchin, 1982). Although the N200 has been described as an index of physical stimulus mismatch, N200's have also been elicited by orthographic, phonological and semantic mismatches (Polich, McCarthy, Wang & Donchin, 1983; Sandquist, Rohbaugh, Syndulko & Lindsely, 1980). Furthermore, the amplitude of the N200 appears to be systematically related to the magnitude of stimulus mismatch, with larger amplitude N200's being elicited by more deviant stimuli (Ford et al., 1976; Kramer et al., 1982). In the present study, the effects of practice and task structure on the N200 will be examined. Since N200 is elicited regardless of the subjects' focus of attention, it is anticipated that neither task structure nor practice will modify the mismatch response.

In a recent study, Hoffman, Simons and Houck (1983) investigated the effects of controlled and automatic processing on components of the ERP. In their study subjects compared either one or four display items to a pair of memory set items. In the VM condition both RT and P300 latency were longer for the larger display set than they were for the smaller set. However, in the practiced CM condition neither RT or P300 latency discriminated between the number of display set items. Unfortunately, the investigators did not record ERPs early in practice so a comparison between the unpracticed and practiced CM condition is impossible. In the present experiment ERPs are recorded both prior to and after the development of the automatic attention response. Thus, ERP components can be examined both as a function of

practice and task structure.

Methods

Subjects

Five male undergraduate students participated in the study. The age of the subjects ranged from 19 to 22. All subjects were right-handed and had either normal or corrected-to-normal vision. Each of the subjects participated in the 12 experimental sessions.

ERP Recording

The EEG was recorded from three midline sites (Fz, Cz & Pz according to the International 10-20 system; Jasper, 1958) and referred to linked mastoids. Two ground electrodes were positioned on the left side of the forehead. Burden Ag-AgCl electrodes, affixed with collodion, were used for scalp and mastoid recording. Beckman bipotential electrodes, affixed with adhesive collars, were placed laterally and supra-orbitally to the right eye to record EOG and this type of electrode was also used for ground recording. Electrode impedances did not exceed 5 kohms/cm.

The EEG and EOG were amplified with Van Gogh model 50000 amplifers (time constant 10 sec and upper half amplitude of 35Hz, 3db/octave roll-off). Both EEG and EOG were sampled for 1800 msec, begining 100 msec prior to stimulus onset. The data were digitized every 10 msec. ERP's were digitally filtered off-line (-3db at 8.8 Hz; 0 db at 20 Hz) prior to statistical analysis.

Stimulus Generation and Data Collection

Stimulus presentation and data acquisition were governed by a PDP 11/40 computer interfaced with an Imlac grapnics processor. Letters and digits which were .45 degrees in height were presented on a Hewlett-Packard CRT which was positioned 65 cm from the subject. Probe letter pairs subtended a visual angle of less than one degree. Single trial EEG and EOG were

performed alone. Performance in the mixed dual-task conditions depended on the instructions given to the subjects. If subjects were instructed to emphasize the CM task, CM performance was maintained at single task levels while VM performance decreased precipitously. However, if VM performance was emphasized, CM and VM tasks were time shared without incurring a performance deficit. It appears that even though resources were unnecessary for CM performance, subjects employed them unless they were instructed not to do so. In the present experiment as well as the study conducted by Hoffman et al. (1983), subjects may have employed resources during the CM task even though they were not needed for successful performance. This would account for the large P300s elicited in the CM conditions in both studies. Thus, the P300s obtained in the practiced CM conditions appear to indicate that subjects employed resources, and not that resources were necessary for automatic processing. A valid test of the hypothesis that automatic processing demands resources would require the recording of ERPs in a dual-task paradigm in which processing priorities could be manipulated.

The differential effects of task structure on the N200 and sustained negativity are noteworthy. The amplitude of the N200 component was larger for the target absent conditions than it was for the target present conditions. N200 amplitudes were also larger during performance with set size four than they were with memory set size one. Thus, N200s were larger when the probes mismatched with the memory set items than when they matched. This mismatch effect was not influenced by the structure of the task or the amount of practice. Therefore, the mismatch processing manifested by the N200 component appears to be performed during both controlled and automatic processing over a wide range of practice levels. This is consistent with previous research which has found that N200 is elicited by stimulus deviance regardless of the focus of subjects' attention (Naatanen et al., 1981;

systematic relationship between P300 amplitude and the cognitive workload of single and dual tasks. In dual-task paradigms, increasing the difficulty of a primary task results in a concomitant decrease in the amplitude of P300s elicited by secondary task events. Conversely, P300s elicited by discrete primary task events increase in amplitude with increases in primary task difficulty (Isreal et al., 1980; Kramer et al., 1983; Natani & Gomer, 1981). The reciprocity of P300 amplitude has been interpreted as evidence for the sensitivity of P300 to resource allocation (Wickens et al., 1983). Changes in P300 amplitude are presumed to reflect the changing resource demands of single and dual tasks. In the present experiment, the decrease in P300 amplitude as a function of increased memory set size in the VM condition may be attributed to an increased need for processing resources while maintaining four items in memory. It is noteworthy that the amplitude of the P300s elicited in the CM condition were unaffected by the size of the memory set. Given that P300 amplitude is a sensitive measure of resource allocation, it appears that increasing the number of items to be maintained in memory in the practiced CM condition does not necessitate the allocation of additional processing resources.

Hoffman, Simons and Houck (1983) have recently argued that the presence of a P300 in the practiced CM condition suggests that CM performance and thereby automatic processing requires resources. Schneider and associates have argued, on the basis of the reduced memory set slope and perfect time sharing of CM with controlled processing tasks, that automatic processing does not require resources (Schneider & Shiffrin, 1977; Schneider & Fisk, 1983). How then are these two sets of results reconciled? In a series of dual-task studies, Schneider and Fisk (1982) paired a VM task with either a CM or VM task. Visual search performance on each of the tasks in the VM dual-task conditions was significantly poorer than when the tasks were

to the completion of stimulus evaluation become automated after extensive practice with consistently mapped stimulus-response pairs. However, as evidenced by the effect of memory set size in the VM condition, practice alone was insufficient to reduce the time required to evaluate task relevant stimuli.

The effect of probability on the amplitude of the P300 component for both CM and VM conditions in session 1 and VM conditions in session 12 was consistent with previously reported findings (Donchin, 1981). P300s elicited by the infrequently occurring events were significantly larger than P300s recorded for the frequent events. However, the probability of target occurrence failed to influence the amplitude of the P300 in the CM condition after extensive practice. This finding cannot be explained as a floor effect in P300 amplitude since the smaller P300s in the VM conditions still manifested the probability effect. The pattern of results suggest that the updating of the mental template, which is manifested by the change in P300 amplitude as a function of probability, does not occur to the same extent when subjects are operating in an automatic processing mode as it does when controlled processing takes place. This is consistent with the proposal that practiced CM performance is not constrained by short term memory limitations (Schneider & Shiffrin, 1977; Fisk & Schneider, 1984).

During VM performance, P300 amplitude decreased from the set size 1 condition to the set size 4 condition. One explanation for this result is the possibility of increased latency variability in the P300s elicited by the probes in the larger memory set condition. However, this explanation is unlikely since the amplitude differences remained even after latency adjusting the single trial waveforms and recomputing the average ERPs. A second, more plausible interpretation concerns the differential resource requirements of the two set size conditions. Previous studies have found a

These results are interesting in that they may signify an increased need for the processing of mismatches when subjects are performing in a controlled processing mode. N200's were elicited by mismatches in both CM and VM conditions while the sustained negativities, which overlapped and extended beyond the temporal epoch of the N200s, distinguished between matches and mismatches only in the VM conditions. Thus, the sustained negativities may reflect additional "controlled processing" of the stimuli.

DISCUSSION

The present experiment examined the effects of extensive practice and task structure on components of the event-related brain potential. The RT results were consistent with previous studies (Shiffrin & Schneider, 1977). Subjects took longer to decide that a target was present when they were required to maintain four items in memory than when they memorized a single item. This effect was obtained in both VM and CM conditions prior to practice and in the VM condition after 23,000 trials of practice on the visual search task. However, in conditions in which subjects were able to consistently map stimuli to responses, the memory set slope was significantly reduced after practice. This reduction in slope in the CM condition is one criterion employed to define automaticity of processing.

The P300 latency results mirrored those obtained with RT. When subjects were unable to consistently map stimuli to responses, the P300s elicited in conditions in which subjects were required to maintain four items in memory were longer than those recorded for the smaller set size. This effect was obtained regardless of the amount of practice received by subjects. The latency of the P300s elicited in the CM conditions after practice did not discriminate between the different memory set sizes. The P300 latency effects obtained in the CM conditions suggest that processes occurring prior

N200 amplitude was replicated in the present study. N200's elicited by the target absent trials were significantly larger than N200s elicited by target present trials ($\underline{F}(1,4)=51.5$, $\underline{p}<.01$). An alternative interpretation to the mismatch hypothesis might be based on the different response requirements for the target present and target absent trials in the present experiment (go-nogo response). However, results of a previous choice RT study with the same subjects indicated that the N200 difference occurred even when the response requirements were the same (Kramer et al., 1983).

N200's were also larger for memory set size 4 than they were for set size 1 ($\underline{F}(1,4)=11.3$, $\underline{p}<.05$). The difference in N200 amplitude as a function of set size may also be attributed to mismatch processing since the probe mismatches with more memory set items with set size 4 than set size 1. The main effects of memory set size and type of response did not interact with the amount of practice or structure of the visual search task. Thus, it appears that the processes reflected by N200 develop relatively quickly and occur in both automatic and controlled processing modes.

<u>Sustained Negativity</u> This frontally negative component overlaps with both N200 and P300 and extends approximately 1200 msec post-stimulus. In the VM conditions in sessions 1 and 12 the sustained negativity was larger for the target absent than the target present trials ($\underline{F}(1,4)=27.9$, $\underline{p}<.01$). The differences in the amplitude of the sustained negativities elicited by the target absent and target present responses during VM performance were further enhanced in conditions in which subjects maintained four items in memory. Sustained negativities elicited in the set size 4 conditions were significantly larger than those elicited when subjects were required to maintain one item in memory ($\underline{F}(1,4)=43.4$, $\underline{p}<.01$). Neither the size of the memory set or the type of response had an effect on the amplitude of the sustained negativities recorded during CM performance.

In session 12 memory set size had a significant effect on P300 latency in the VM conditions. P300 latencies were longer for set size 4 than they were for set size 1 ($\underline{F}(1,4)=16.3$, $\underline{p}<.01$). The size of the memory set did not have a significant effect on P300 latency in the CM conditions. Thus, the pattern of P300 latencies elicited by the probe stimuli is consistent with the RT results. RT and P300 latency distinguished between memory set conditions in the first experimental session and for the VM conditions in session 12. In the practiced CM condition neither RT or P300 latency was influenced by the size of the memory set. These findings suggest that processes occurring prior to the completion of stimulus evaluation are becoming automated during CM practice.

In both experimental sessions, the P300s elicited in the target present conditions were shorter in latency than the P300s recorded in the target absent conditions ($\underline{F}(1,4)=8.0$, $\underline{p}<.05$). The increased latency of the P300s in the target absent conditions is consistent with previously reported research in which subjects performed visual search tasks with a CRT response (Brookhuis et al., 1983; Hoffman et al., 1983). In the target absent conditions P300 latency was shorter when the targets occurred frequently than when they occurred infrequently ($\underline{F}(1,4)=9.4$, $\underline{p}<.05$). The set size slope for the VM target absent trials in session 1 was significantly larger (50 vs 34 msec) than the slope for the target present conditions, possibly indicating a self terminating memory search process ($\underline{F}(1,4)=10.1$, $\underline{p}<.05$). For the CM conditions in both sessions and the VM condition in session 12, the slopes of the target present and target absent conditions were not significantly different. Similar slopes in target present and target absent conditions have been used as evidence for exhaustive search through memory (Sternberg, 1969).

<u>N200 Amplitude</u> The commonly observed effect of stimulus mismatch on

The amplitude of the P300 was influenced by the type of response ($F(1,4)=44.5$, $p<.01$). P300s elicited in the target present conditions were larger than P300s recorded during the target absent conditions. Although the presence or absence of the target was confounded with the go/no-go response task, results from a choice RT study with the same subjects indicated that P300 amplitude is larger for target present trials even if an overt response is required for the target absent trials (Kramer, Fisk & Schneider, 1983). Thus, the larger P300 amplitude in the target present conditions cannot be attributed to differences in response requirements between the target present and target absent conditions.

Insert Figure 8 about here

P300 Latency Figure 8 presents the P300 latencies for all experimental conditions in sessions 1 and 12. A comparison of the top panel of Figure 8 with Figure 1 reveals the similarity in the patterns of the P300 latency and RT data. The most noteworthy effect was the significant three-way interaction between session, task structure and memory set size ($F(1,4)=53.9$, $p<.01$), indicating that memory set size did not influence P300 latency in the practiced CM condition.

In session 1 the latency of the P300s elicited by the probe stimuli was longer when subjects were maintaining four items in memory than when they were required to memorize a single item ($F(1,4)=121.1$, $p<.01$). P300 latencies were significantly shorter when subjects were performing the visual search task in the CM conditions than when they were responding in the VM conditions ($F(1,4)=18.4$, $p<.01$). The differences between P300 latencies as a function of memory set size were larger in the VM than in the CM conditions ($F(1,4)=40.8$, $p<.01$).

$p<.01$). This effect is illustrated in Figure 7. In session 1 P300s elicited by low probability stimuli (either targets or distractors) were significantly larger than P300s elicited by high probability stimuli ($F(1,4)= 18.6$, $p<.01$). This finding is consistent with previously published studies which have investigated the effects of probability on P300 amplitude (see Donchin, Ritter & McCallum, 1978). However, in session 12 P300 amplitudes were larger for the low probability than the high probability stimuli for the VM but not the CM condition ($F(1,4)=13.7$, $p<.05$). Thus, it would appear that the memory updating which is manifested by the P300 is reduced after the development of the automatic attention response. Furthermore, practice alone is insufficient to reduce the updating process since the effect is not diminished in the VM conditions. An alternative interpretation of the lack of a probability effect in the practiced CM condition is a floor effect in P300 amplitude. However, this interpretation appears untenable in light of the significant probability effect in the VM condition in session 12 (see figure 7).

The P300s elicited by probe stimuli when subjects were maintaining four items in memory were significantly smaller than P300s for set size 1 ($F(1,4)= 13.9$, $p<.05$). Furthermore, a significant interaction between task structure and memory set size was obtained ($F(1,4)=17.4$, $p<.01$). P300s elicited by the VM stimuli decreased in amplitude from set size 1 to set size 4 while the amplitude of the P300s elicited in the CM conditions was not influenced by the size of the memory set. The decreased amplitude in the VM condition may indicate an increased need for resources to maintain four items in memory (Isreal et al., 1980; Kramer et al., 1983). Presumably, the automatic processing which takes place in the CM conditions is not influenced by the size of the memory set, thereby obviating the requirement for additional resources during performance with the larger memory set size.

The latency adjustment procedure yields latency adjusted single-trial waveforms which can be averaged according to experimental conditions and submitted to the PCA. The latency adjustment procedure also determines component peak latencies for each subject in each condition. These latency values can then be analyzed much in the same manner as RT.

Thus, two separate PCA's were performed: one on unadjusted, the other on latency adjusted ERPs. The component scores obtained from the first PCA were used to assess the effect of experimental manipulations on ERP components other than the P300. The second PCA was employed to evaluate the effect of experimental factors on the P300.

Insert Figure 6 about here

The PCAs revealed the existence of several components in the data. Figure 6 shows the Varimax rotated component loadings for the first three factors extracted by the PCA of the unadjusted average waveforms. The three PCA components correspond to the temporal ranges and electrode distributions of the ERP components described on the basis of a visual inspection of the data. This is not to say that the components derived in the PCA are synonymous with the ERP components, only that the experimental variance represented by the PCA components presents a pattern of results consistent with the ERP components.

Insert Figure 7 about here

P300 Amplitude The prediction of a reduced P300 probability effect in the practiced CM condition was supported by a significant three-way interaction between session, task structure and probability ($F(1,4)=76.6$,

especially for the larger memory set size. It is difficult to tell from visual inspection of the data whether this component is distinct from the earlier negativity or just a continuation of it.

Insert Figures 2,3,4 & 5 about here

Principal Components Analysis The effects of the experimental variables on the ERP components were assessed with a Principal Component Analysis (PCA) technique (Coles, Gratton, Kramer & Miller, in press; Donchin & Heffley, 1979). Two separate PCAs were performed on the data. In the first procedure the components of the ERP were derived and analyzed by applying the PCA technique to the averaged raw data. The data base submitted to the PCA consisted of 480 ERPs (5 subjects x 2 sessions X 2 memory sets x 2 probabilities x 3 electrodes x 2 task structures x 2 target states), each composed of 180 time points. The PCA was performed on the covariance matrix obtained by computing the covariance between the voltages recorded at each pair of time points over the entire data set. The component scores computed from a linear combination of time points by loading coefficients were then subjected to a repeated measures ANOVA.

Visual inspection of the waveforms presented in Figures 2 through 5 reveals a high degree of latency variability in the P300 component. This variability in latency makes it difficult to interpret the results of the PCA. A latency adjustment procedure (Woody, 1967) was therefore employed to create a data set in which P300 latency was less variable. The latency adjustment procedure uses a lagged cross correlation algorithm to detect and align the desired component so as to eliminate the latency jitter across single trials. This procedure helps reduce the confounding of changes in amplitude with the variability in the latency of a specific ERP component.

$p<.05$). A significant interaction between memory set size and task structure was also obtained such that higher error rates were associated with the VM condition, especially for the larger set size (1.2% vs 3.4%; $F(1,4)=7.3$, $p<.05$). In the VM conditions subjects made more errors when maintaining four items in memory in session 1 than they did in session 12 (4.6% vs 2.2%; $F(1,4)=9.4$, $p<.05$). Since the higher error rates were associated with the longer RTs in all but two VM conditions these data suggest that subjects were not generally trading speed for accuracy when performing the visual search task.

Event-Related Potentials

Figures 2 and 3 present the average ERPs elicited by the probe stimuli in all of the CM conditions in sessions 1 and 12, respectively. The same information is presented for the VM conditions in Figures 4 and 5. Several deflections in the waveforms are noteworthy. The average waveforms possess a positive going deflection which occurs approximately 450 msec following the presentation of a stimulus and is maximum in amplitude at the parietal electrode. This component appears to be larger for the target present trials than it is for the target absent trials. Probability also seems to have an effect on this positive deflection with larger amplitude components being elicited by the low probability events. The latency range, electrode distribution and sensitivity to the manipulation of probability are consistent with the criteria employed in the definition of the P300 (Donchin, Kramer & Wickens, 1982). A frontally maximal, negative going deflection which occurs approximately 350 msec after the presentation of a stimulus is also visible in the average waveforms. This component appears to be larger in the target absent conditions than it is in the target present trials. Another relatively slow component which occurs subsequent to the first negative deflection can be found in the target absent conditions,

---

Insert Figure 1 about here

In session 12, after subjects had received over 23,000 trials of practice, the size of the memory set still produced a significant effect in the VM condition ($\underline{F}(1,4)=103.4$, $\underline{p}<.01$). RTs were longer for set size 4 than they were for set size 1. In the CM conditions, the size of the memory set did not have a significant effect on RT. No other main effects or interactions were statistically significant. The pattern of RTs produced in the CM and VM conditions is consistent with previous findings (Schneider & Shiffrin, 1977). Even extensive practice does not improve performance when subjects are unable to consistently map stimuli to responses (VM condition). However, when subjects are able to consistently map stimuli to responses, practice ultimately leads to an automaticity of processing as suggested by the failure of memory set size to influence RT.

The RTs elicited in the VM condition during session 12 were actually longer than the RTs produced during session 1, especially for conditions in which four memory set items were presented ($\underline{F}(1,4)=18.5$, $\underline{p}<.01$). A possible explanation for the increased RT in the VM condition in session 12 is the difference in error rates between sessions. Subjects made significantly fewer errors in session 12 than they did in session 1, particularly for the larger memory set condition. Thus, the longer RT in session 12 may be the result of a speed-accuracy tradeoff.

The mean error rate across all of the experimental conditions in sessions 1 and 12 was 1.4%. Subjects made significantly more errors when they were required to maintain four items in memory (2.1% vs .75%) than they did when only one memory set item was presented ($\underline{F}(1,4)=13.9$, $\underline{p}<.05$). VM trials resulted in more errors than CM trials (2.3% vs .45%; $\underline{F}(1,4)=8.6$,

the sessions. ERPs were recorded in the first and twelfth sessions.

The temporal sequence of each set of trials in the first and twelfth sessions was as follows: A memory set of one or four elements was presented for 10 sec and was followed by a blank screen for 1000 msec. The 30 probe trials that followed the presentation of the memory set began with the presentation of two elements for 200 msec. ISI's were 1800 msec. In sessions two through eleven the 30 probe trials began with a 100 msec presentation of the two elements. ISI's were 900 msec. The stimulus presentation epoch and ISI's were shortened in these sessions in order to decrease the time required for the acquisition of the automatic response.

Results

Reaction Time and Error Rates

RT was defined as the interval between the appearance of the probe stimuli and the subjects key-press indicating that a memory set item was presented. Subjects were to respond only when a memory set item was displayed (go/no-go task). Figure 1 presents the average RTs for each of the experimental conditions in sessions 1 and 12. The most noteworthy effect was the significant three-way interaction among session, task structure and memory set size ($F(1,4)=91.6$, $p<.01$), indicating that memory set size did not have an effect on RT in the CM conditions in session 12.

In session 1, subjects took longer to decide if a target was presented with set size 4 than they did with set size 1 ($F(1,14)=31.4$, $p<.01$). Subjects also responded more quickly when stimuli could be consistently mapped to responses (CM condition) than they did when they were unable to consistently map the stimuli to responses ($F(1,4)=44.5$, $p<.01$). The difference in the RTs as a function of the size of the memory set was larger in the VM than in the CM condition ($F(1,4)=61.7$, $p<.01$).

monitored on-line using a GT-44 display. Digitized single trial data were stored on magnetic tape for later analysis.

Evaluation of each EOG record for saccades and blinks was conducted off-line by calculating its variance and comparing this to a preset criterion for acceptance. Single trial EEG containing unacceptable EOG was discarded prior to statistical analysis.

Procedure

The subjects' task was to decide if a letter or digit presented on the CRT belonged to a previously memorized set of elements. Each set of 30 trials began with a 10 sec presentation of a memory set of letters or digits. Memory sets included either one or four elements. In the thirty trials that followed the presentation of each memory set, the subjects' task was to press a button if a memory set item (target) was present (go/no-go task). Each of the trials contained two items, either a target and a distractor or two distractors.

Three variables were orthogonally manipulated in a within subjects, factorial design. These variables included the number of memory set items (one or four), the task structure (CM or VM), and the probability of occurence of a memory set item (.20 or .80). In the CM condition targets were always selected from one category (numbers 1 to 9) while distractors were chosen from another category (letters A to I). In the VM condition both the targets and distractors were choosen from the same category (letters A to I). Targets and distractors exchanged roles over trials in the VM conditions. Twelve sessions of 1920 trials were run with each of five subjects. The structure of the task, the probability of occurrence of a memory set item and the number of memory set items served as blocking factors. Subjects performed eight blocks of 240 trials in each of the 12 experimental sessions. RT's and accuracy measures were obtained in all of

Squires et al., 1977).

The sustained negativity was also affected by both response type and memory set size with larger components being elicited by the target absent and set size four conditions. However, the effect of these variables on the sustained negativity occurred only in the VM condition. Thus, although both automatic and controlled processing tasks displayed an N200 mismatch effect, only the controlled processing task produced a sustained negativity mismatch effect. The difference in sustained negativity as a function of stimulus mismatch may signify an increased need for the processing of mismatches when subjects are performing in the controlled processing mode.

ERPs recorded in conjunction with RT and error rate have provided insights into the processing underlying performance in automatic and controlled processing tasks. The negative components of the ERP have provided information concerning the processing of mismatches which is not easily obtained with more traditional measures. The failure to find a significant P300 probability effect in the practiced CM conditions suggests that automatic processing is not constrained by the same short term memory limitations as controlled processing. Further research, in more complex dual-task paradigms, will be necessary to examine the effects of priorities and task structure on the resource allocation between CM and VM tasks.

Figure Captions

Figure 1. The average RT's for each of the experimental conditions in sessions 1 and 12.

Figure 2. The average ERP's elicited by the probe stimuli in all of the CM conditions in session 1.

Figure 3. The average ERP's elicited by the probe stimuli in all of the CM conditions in session 12.

Figure 4. The average ERP's elicited by the probe stimuli in all of the VM conditions in session 1.

Figure 5. The average ERP's elicited by the probe stimuli in all of the VM conditions in session 12.

Figure 6. The component loadings for the first three components extracted from a Principal Components Analysis of the average ERPs.

Figure 7. Average P300 amplitude as a function of session, task structure and target probability.

Figure 8. The P300 latencies for all of the experimental conditions in sessions 1 and 12. Average P300 latency values were obtained by averaging single trial latencies obtained from a Woody latency adjustment procedure.

References

Adam, N. & Collins, G.I. (1978). Late components of the visual evoked potential to search in short-term memory. Electroencephalography and Clinical Neurophysiology, 44, 147-156.

Brookhuis, K.A., Mulder, G., Mulder, L.J.M. & Gloerich, A.B.M. (1983). The P300 complex as an index of information processing: The effects of response probability. Biological Psychology, 17, 277-296.

Coles, M.G.H., Gratton, G., Kramer, A. & Miller, G. (in press). Principles of signal acquisition and analysis. In M.G.H. Coles, E. Donchin & S.W. Porges (Eds.), Psychophysiology: Systems, processes and applications. New York: Guilford Press.

Donchin, E. (1981). Surprise! ... Surprise? Psychophysiology, 18, 493-515.

Donchin, E. & Heffley, E. (1979). Multivariate analysis of event related potential data: A tutorial review. In D. Otto (Ed.), Multidisciplinary Perspectives in Event Related Potential Research. (EPA 600/9-77-043). Washington, D.C.: U.S Government Printing Office.

Donchin, E. & Isreal, J.B. (1980). Event-related brain potentials and psychological theory. In H.H. Kornhuber & L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavior and Clinical Use. Amsterdam: Elsevier-North Holland.

Donchin, E., Kramer, A. & Wickens, C. (1982). Probing the cognitive infrastructure with event-related brain potentials. Proceedings of the AIAA Workshop on Flight Testing to Identify Pilot Workload

and Pilot Dynamics. Washington, D.C.: U.S. Government Printing Office.

Donchin, E., McCarthy, G., Kutas, M. & Ritter, W. (1984). Event-related brain potentials in the study of consciousness. In R.J. Davidson, G.E. Schwartz & D. Shapiro (Ed.), Consciousness and Self-Regulation, Vol 3. New York: Plenum Publishing Corporation.

Donchin, E., Ritter, W. & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting & S. Koslow (Eds.), Brain Event-Related Potentials in Man. New York: Academic Press.

Duncan-Johnson, C.C. (1981). P300 latency: A new metric for information processing. Psychophysiology, 18, 207-215.

Duncan-Johnson, C.C. & Donchin, E. (1977). On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 14, 456-467.

Duncan-Johnson, C.C. & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. Biological Psychology, 14, 1-52.

Fisk, A.D. & Schneider, W. (1983). Category and word search: Generalizing search principles to complex processing. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 177-195.

Fisk, A.D. & Schneider, W. (1984). Memory as a function of attention, level of processing, and automatization. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 181-197.

Ford, J.M., Mohs, R.C., Pfefferbaum, A. & Kopell, B.S. (1980). On the utility of P300 latency and reaction time for studying cognitive processes. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain:

Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press.

Ford, J.M., Pfefferbaum, A. & Kopell, B.S. (1982). Event-related potentials to a change of pace in a visual sequence. Psychophysiolcgy, 19, 173-177.

Ford, J.M., Roth, T.R. & Kopell, B.S. (1976). Auditory evoked potentials to unpredictable shifts in pitch. Psychophysiology, 13, 32-39.

Ford, J.M., Roth, W.T., Mohs, R.C., Hopkins, L.F. & Kopell, B.S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 47, 450-454.

Gomer, F.E., Spicuzza, R.J. & O'Donnell, R.D. (1976). Evoked correlates of visual item recognition during memory scanning tasks. Physiological Psychology, 4, 61-65.

Hoffman, J.E., Simmons, R.F. & Houck, M.R. (1983). Event-related potentials during controlled and automatic target detection. Psychophysiology, 20, 625-632.

Isreal, J.B., Wickens, C.D., Chesney, G.L. & Donchin, E. (1980). The event-related brain potential as an index of display monitoring workload. Human Factors, 22, 211-224.

Jasper, H.H. (1958). The ten twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, 10, 371-375.

Karis, D., Fabiani, M. & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. Cognitive Psychology, 16, 177-216.

Klinke, R., Fruhsterter, H. & Finkenzeller, P. (1968). Evoked

responses as a function of external and stored information. Electroencephalography and Clinical Neurophysiology, 25, 119-122.

Kramer, A.F., Fisk, A. & Schneider, W. (1983). P300, consistency and visual search. Psychophysiology, 20, 453-454.

Kramer, A., Ross, W. & Donchin, E. (1982). Individual differences in human information processing: A chronometric analysis of the role of orthographic and phonological cues in a non-lexical decision task. Psychophysiology, 19, 330-331.

Kramer, A.F., Wickens, C.D. & Donchin, E. (1983). An analysis of the processing requirements of a complex perceptual-motor task. Human Factors, 25, 597-621.

Magliero, A., Bashore, T.R., Coles, M.G.H. & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21, 171-186.

McCarthy, G. & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. Science, 211, 77-80.

Naatanen, R. (1982). Processing negativity: An evoked potential reflection of selective attention. Psychological Bulletin, 92, 605-640.

Naatanen, R. & Gaillard, A.W.K. (1983). The orienting reflex and the N2 deflection of the ERP. In A.W.K. Gaillard and W. Ritter (Eds.), Tutorials in Event Related Potential Research: Endogenous Components. Amsterdam: North Holland Publishing Company.

Naatanen, R., Gaillard, A.W.K. & Varey, C.A. (1981). Attention effects on auditory EPs as a function of inter-stimulus interval. Biological Psychology, 13, 173-187.

Naatanen, R., Simpson, M. & Loveless, N.E. (1982). Stimulus deviance and evoked potentials. Biological Psychology, 14, 53-98.

Natani, K. & Gomer, F.E. (1981). Electrocortical activity and operator workload: A comparison of changes in the electroencephalogram and in event-related brain potentials. McDonnell Douglas Technical Report, MDC E2427. McDonnell Douglas Astronautics Company, St. Louis, Missouri.

Polich, J., McCarthy, G., Wang, W.S. & Donchin, E. When words collide: Orthographic and phonological interference during word processing. Biological Psychology, 16, 155-180.

Sandquist, T.F., Rohbaugh, J., Syndulko, K. & Lindsely, D.B. (1980). An event-related potential analysis of coding processes in human memory. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press.

Schneider, W. & Fisk, A.D. (1982). Concurrent automatic and controlled visual search: Can processing occur without resource cost? Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 261-277.

Schneider, W. & Fisk, A.D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. Perception and Psychophysics, 31, 160-168.

Schneider, W. & Fisk, A.D. (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 1-15.

Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human

information processing: I. Detection, search and attention. Psychological Review, 84, 1-66.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review, 84, 127-190.

Sokolov, E.W. (1969). The modeling properties of the nervous system. In I. Maltzman & K. Cole (Eds.), Handbook of Contemporary Soviet Psychology. New York: Basic Books.

Squires, N.K., Donchin, E., Herning, R.I. & McCarthy, G. (1977). On the influence of task relevance and stimulus probability on event-related brain potential components. Electroencephalography and Clinical Neurophysiology, 42, 1-14.

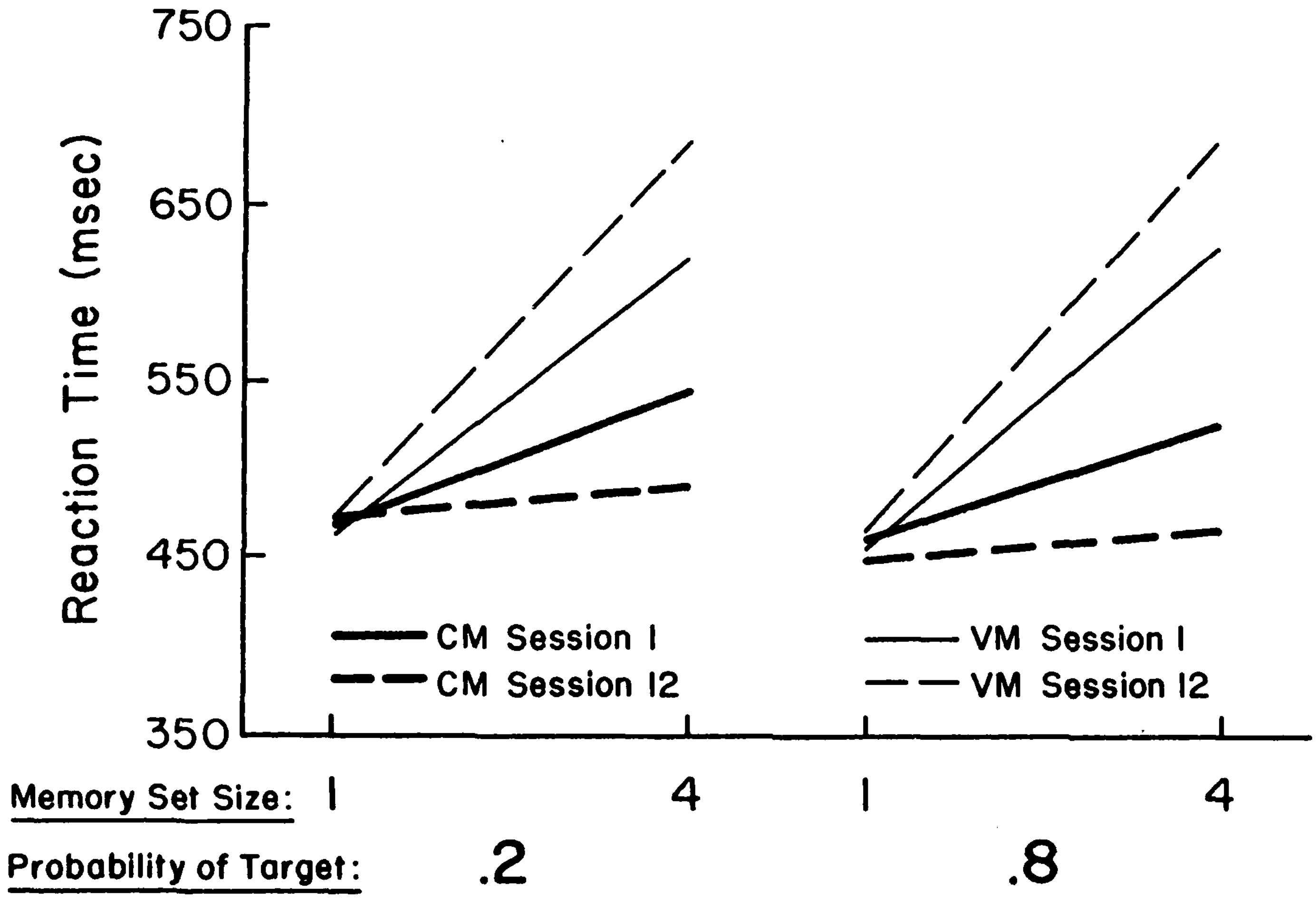
Squires, N.K., Donchin, E., Squires, K.C. & Grossberg, S. (1979). Bisensory stimulation: Infering decision related processes from the P300 component. Journal of Experimental Psychology: Human Perception and Performance, 3, 299-315.

Sternberg, S. (1966). High speed scanning in human memory. Science, 153, 652-654.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta Psychologica, 30, 276-315.

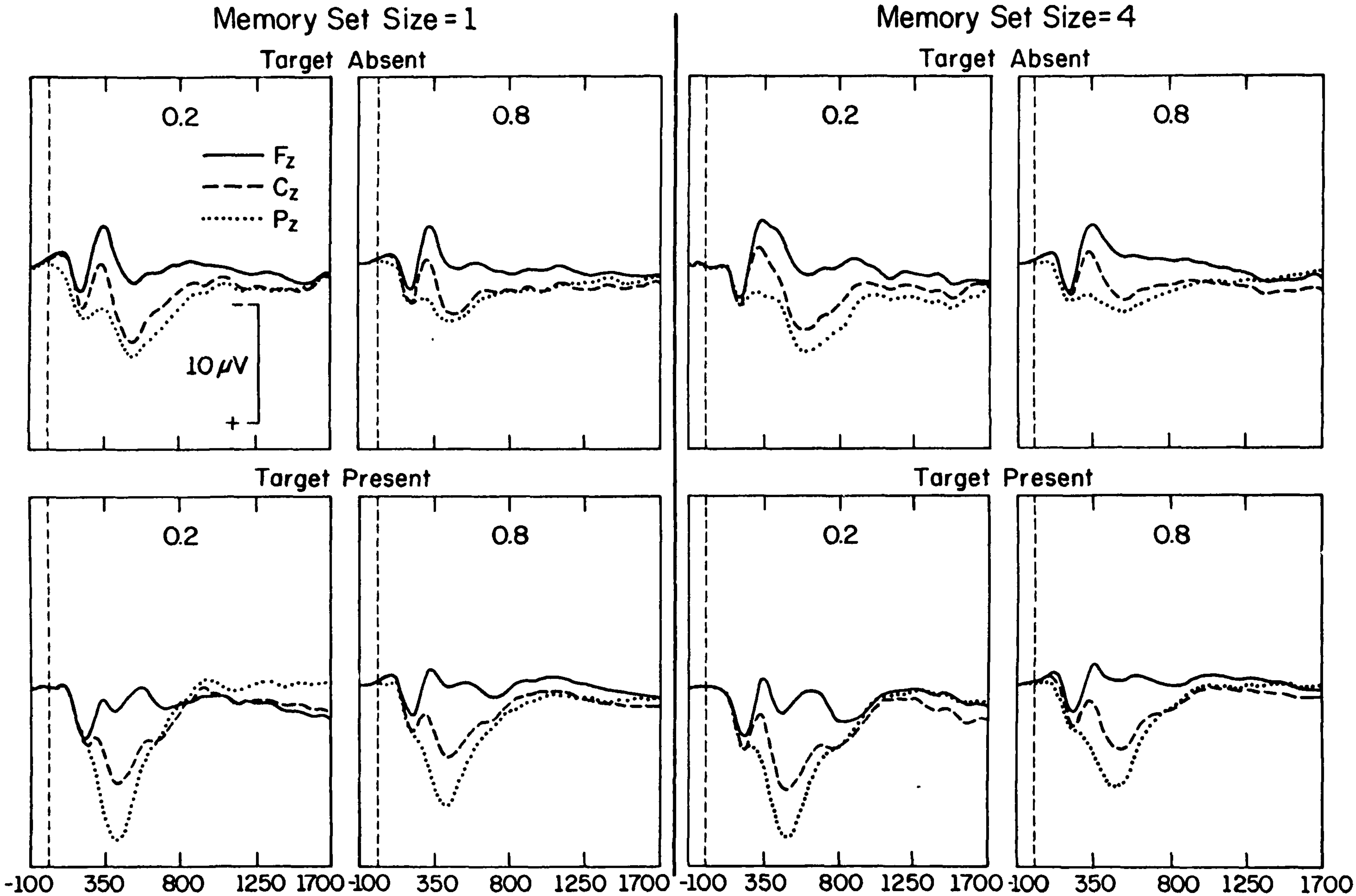
Wickens, C., Kramer, A., Vanasse, L. & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resources. Science, 221, 1080-1082.

Woody, C.D. (1967). Characterization of an adaptive filter for the analysis of varaible latency neuroelectric signals. Medical Biological Engineering, 5, 539-553.

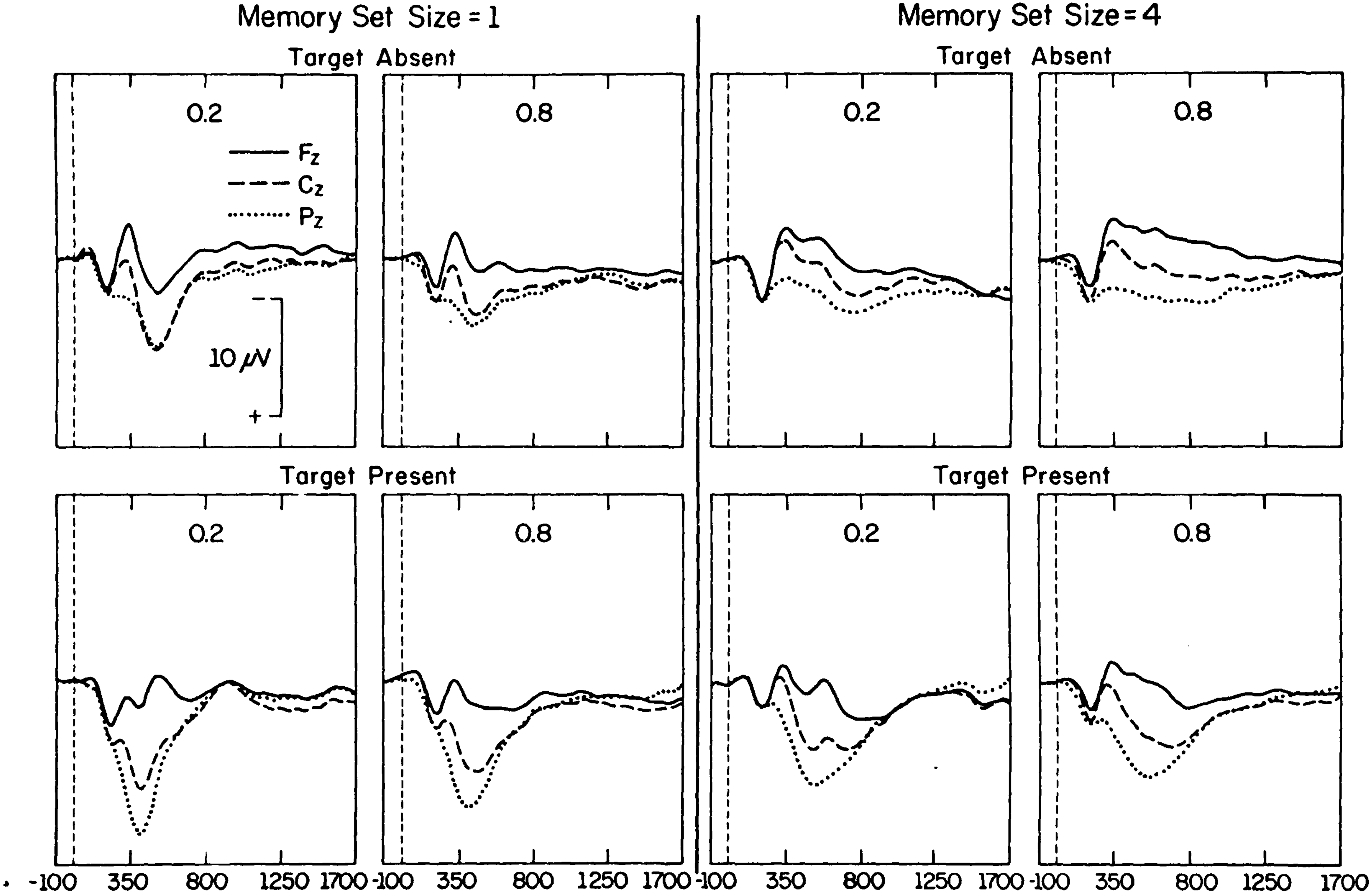


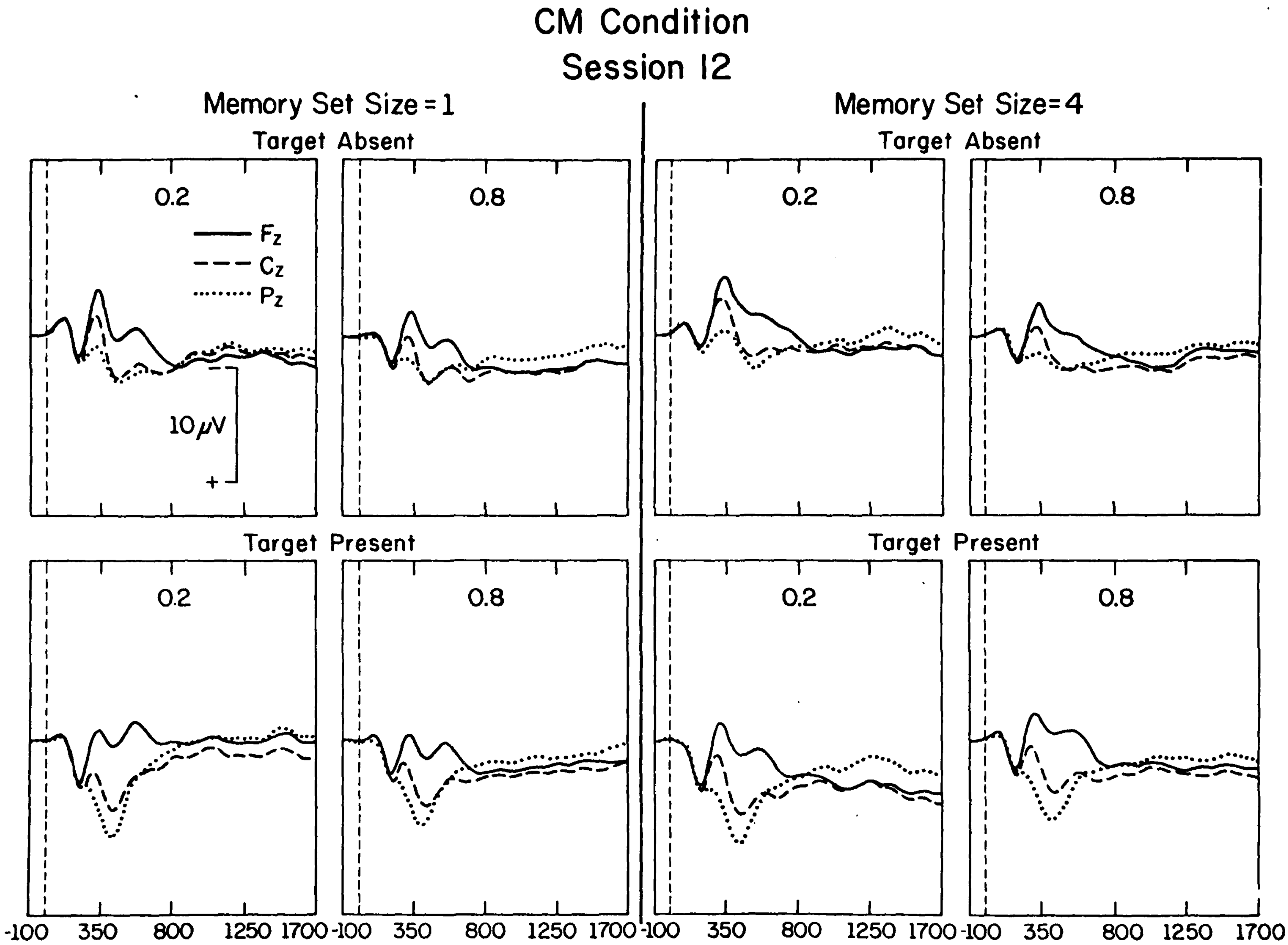
Reaction Time (msec)
750
650
550
450
350
CM Session 1
CM Session 12
VM Session 1
VM Session 12
Memory Set Size: 1 4 1 4
Probability of Target: .2 .8

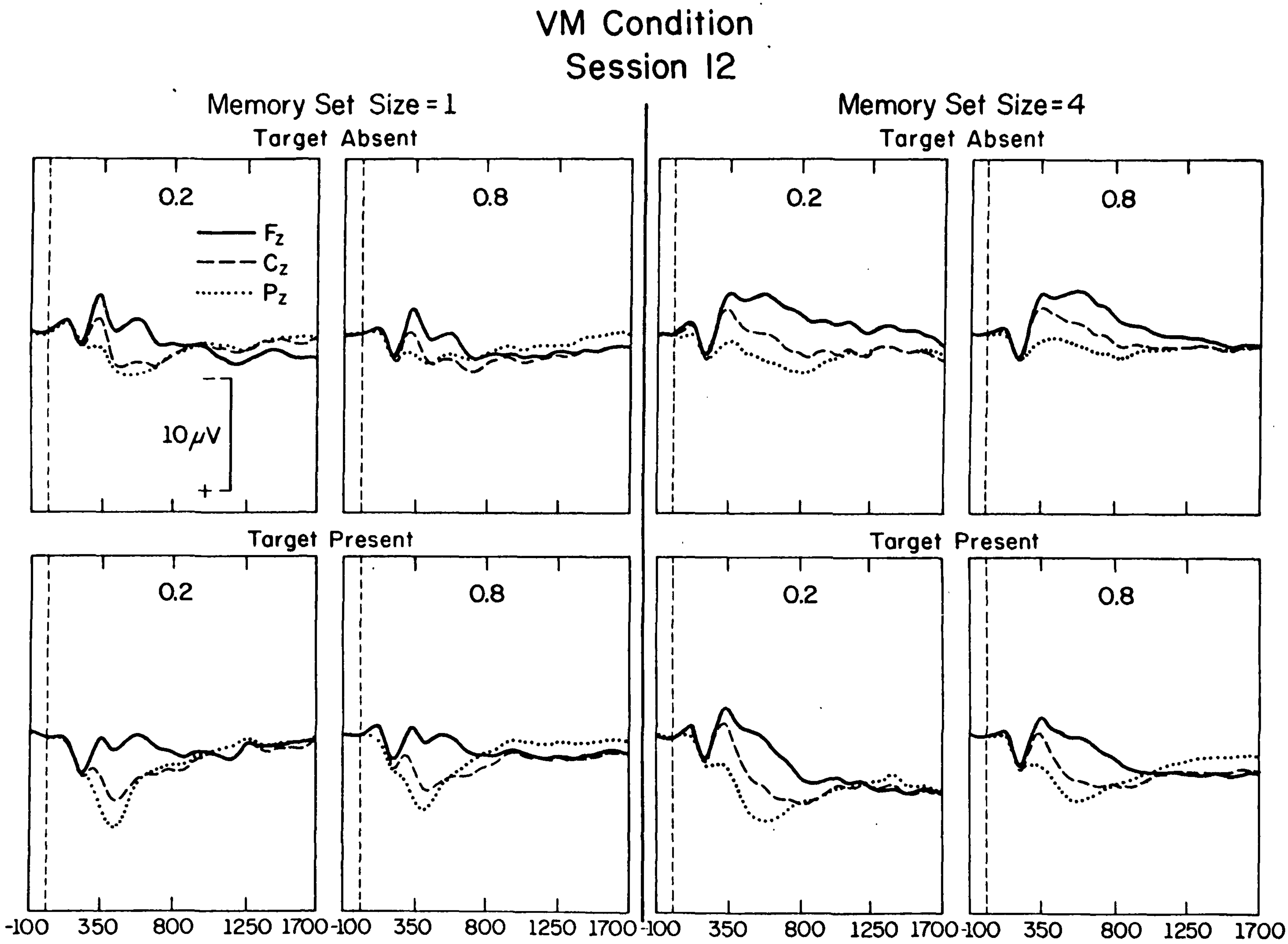
CM Condition
Session 1



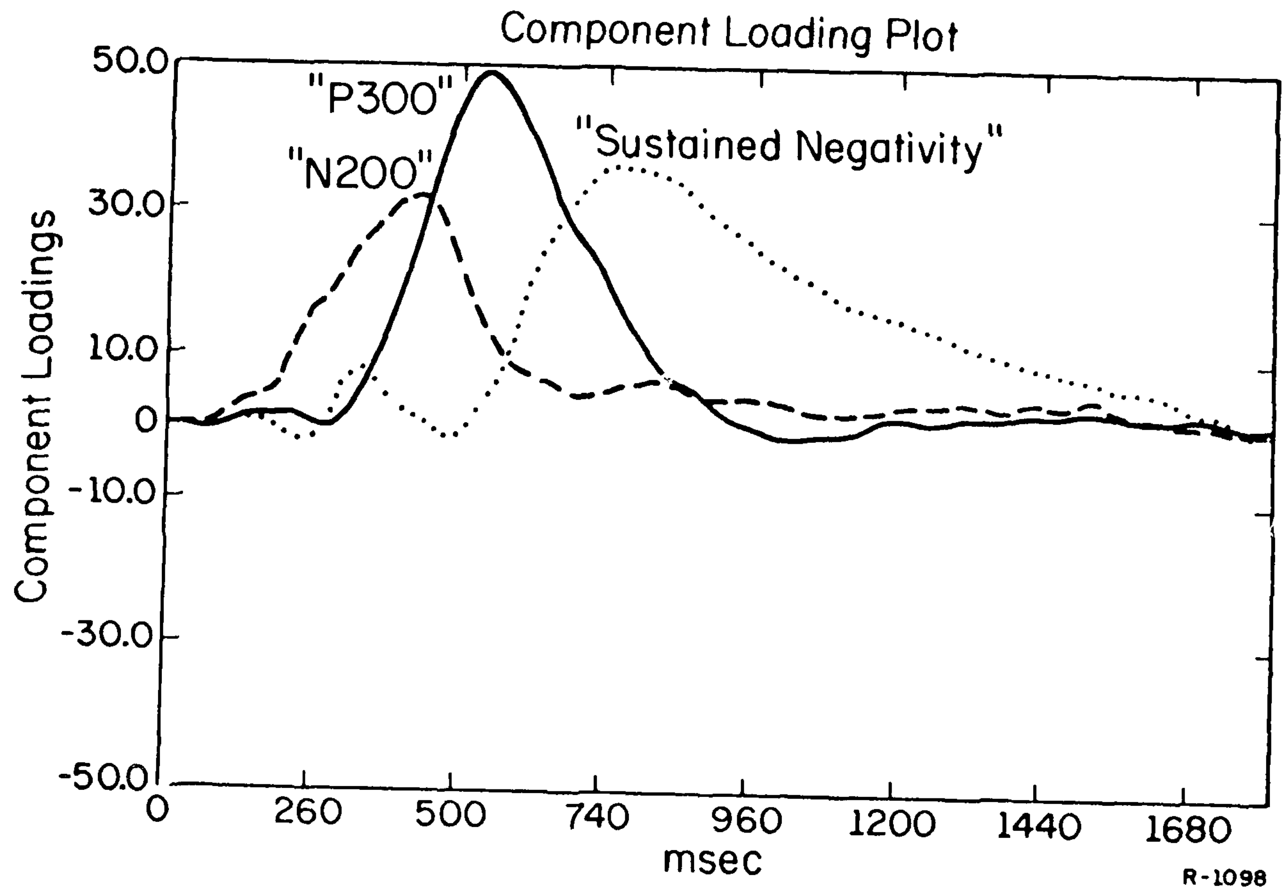
VM Condition
Session 1




CM Condition
Session 12
Memory Set Size = 1
Memory Set Size = 4
Target Absent
Target Present
0.2
0.8
Fz
Cz
Pz
10μV
+
-100
350
800
1250
1700

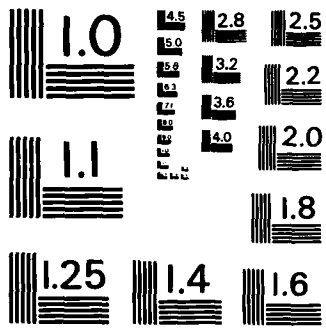


VM Condition
Session 12
Memory Set Size = 1
Target Absent
Target Present
Memory Set Size = 4
Target Absent
Target Present
0.2
0.8
F_z
C_z
P_z
10 μV
+
-100
350
800
1250
1700

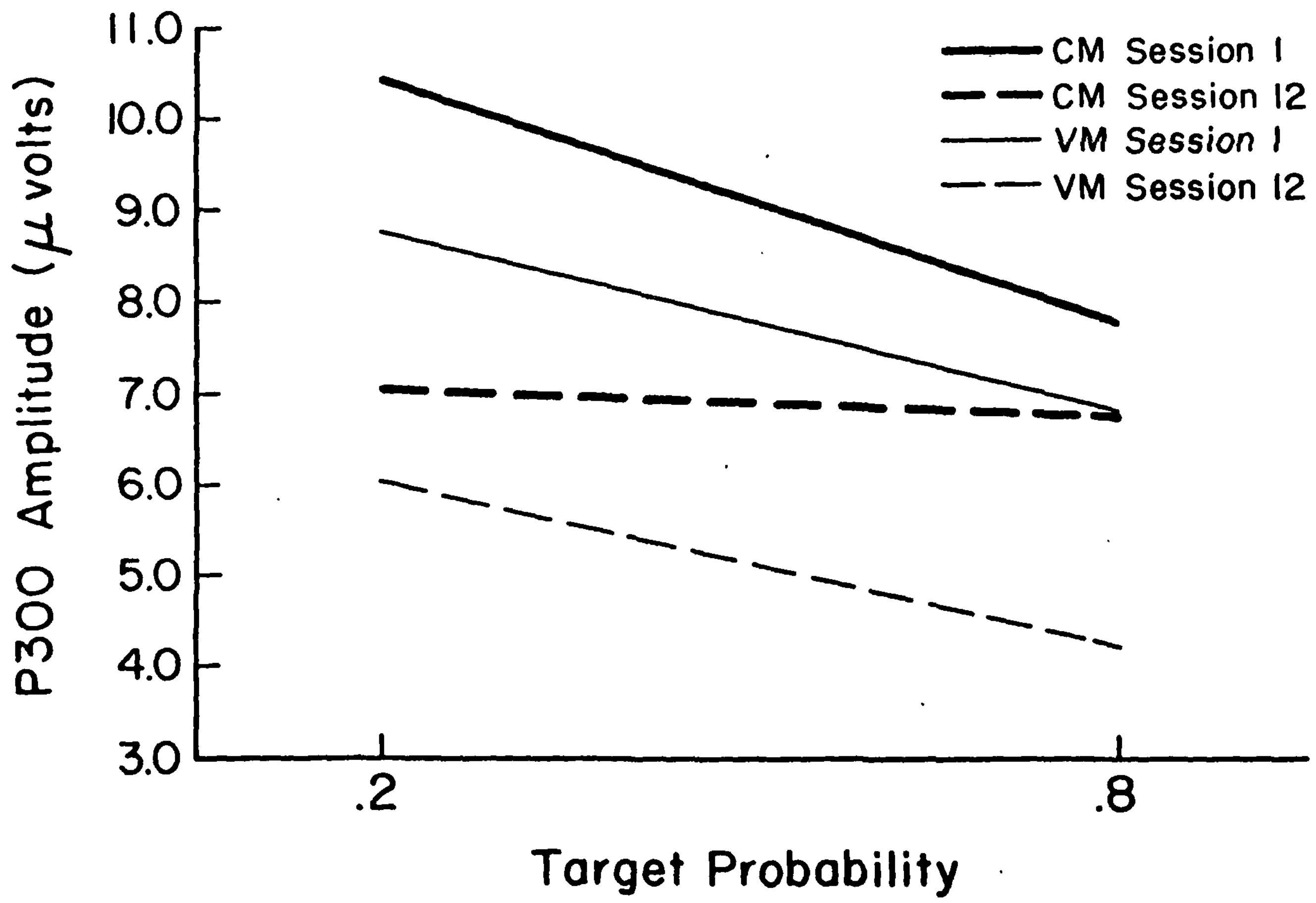

Component Loading Plot
"P300"
"N200"
"Sustained Negativity"
Component Loadings
50.0
30.0
10.0
0
-10.0
-30.0
-50.0
0
260
500
740
960
1200
1440
1680
msec
R-1098

AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 7/9

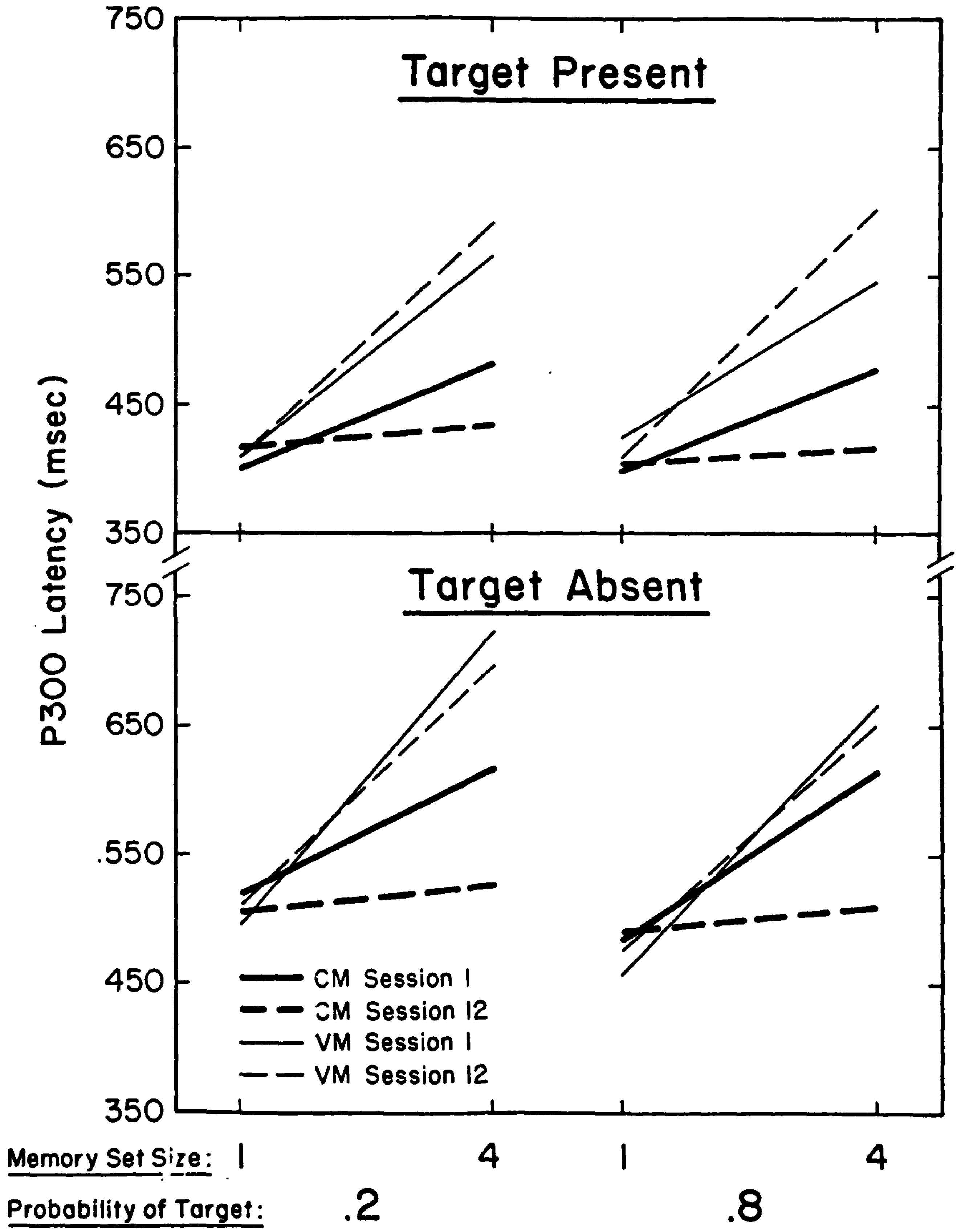
UNCLASSIFIED 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



CM Session 1
CM Session 12
VM Session 1
VM Session 12
P300 Amplitude (μ volts)
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
.2
.8
Target Probability


Target Present
Target Absent
P300 Latency (msec)
750
650
550
450
350
CM Session 1
CM Session 12
VM Session 1
VM Session 12
Memory Set Size: 1 4 1 4
Probability of Target: .2 .8

The Processing of Stimulus Properties: Evidence for Dual-Task Integrality

Arthur F. Kramer, Christopher D. Wickens and Emanuel Donchin
Department of Psychology
University of Illinois
Champaign, Illinois

Running Head: Evidence for Dual-Task Integrality

Abstract

Research on dual-task performance has been concerned with delineating the antecedent conditions which lead to dual-task decrements. Capacity models of attention, which propose that a hypothetical resource structure underlies performance, have been employed as predictive devices. These models predict that tasks which require different processing resources can be more successfully time shared than tasks which require common resources. We suggest that dual-task decrements can be avoided even when the same resources are required by both tasks, by designing the tasks so that the processing demands can be integrated. The conditions under which such dual-task integrality can be fostered were assessed in a study in which we manipulated four factors likely to influence the integrality between tasks: inter-task redundancy, the spatial proximity of primary and secondary task displays, the degree to which primary and secondary task displays constitute a single object, and the resource demands of the two tasks. The resource allocation policy associated with these integrated dual-task pairs is inferred from changes in the amplitude of the P300 component of the Event-Related Brain Potential (ERP). Twelve subjects participated in three experimental sessions in which they performed both single and dual-tasks. The primary task was a pursuit step tracking task. The secondary tasks required subjects to discriminate between different intensities or different spatial positions of a stimulus. Task pairs which required the processing of different properties of the same object resulted in better performance than task pairs which required the processing of different objects. Furthermore, these same object task pairs led to a positive relation between primary task difficulty and the resources allocated to secondary task stimuli. Inter-task redundancy and the physical proximity of task displays produced similar effects of reduced magnitude. The results are discussed in terms of a model of dual-task integrality.

The Processing of Stimulus Properties: Evidence for Dual-Task Integrality

Arthur F. Kramer, Christopher Wickens and Emanuel Donchin

The concurrent processing of information relevant to several tasks has been addressed in the psychological literature from the early writings of James to contemporary investigations of dual-task performance in complex, operational environments (James, 1890). Substantial theoretical and empirical effort has been expended in mapping the conditions under which the demands imposed by tasks performed concurrently interact so that performance on one, or both, tasks degrades. Navon and Gopher (1979) proposed that these conditions are related to the resource structure underlying a dual-task combination. The extent of dual-task interference is predicted on the basis of the overlap of processing resources. Tasks which require separate processing resources will be more sucessfully time shared than tasks which require common processing resources. Wickens (1984) proposed a Multiple Resource Model according to which processing resources may be represented by three dimensions: stages of processing, modalities of processing and codes of processing. This theoretical conceptualization of the processing structure of dual-task performance has received considerable empirical support. Attempts to perform concurrently tasks which require processing resources from the same modalities, codes or stages of processing generally result in larger decrements in performance than does the concurrent performance of tasks which require resources from different structures (Alwitt, 1981; Isreal, 1980; North, 1977; Trumbo, Noble and Swink, 1967; Wickens and Kessel, 1979; Wickens, Sandry and Vidulich, 1983).

This paper is concerned with the availability of cooperative processing strategies that make it possible for the processing routines necessary to accomplish one task to be used in the processing of another task. This emphasis on cooperative processing strategies is in marked contrast with the common approach to the analysis of the performance of dual-tasks. In general, investigators have assumed that the processing necessary to meet the performance criteria for one of the tasks is either independent of, or at odds with, the processing required for meeting the criteria for the other task. That is, the emphasis has always been on the competition for resources between tasks. In this paper we focus on the cooperation between tasks. Specifically, we propose that the stimulus ensembles associated with different tasks can, on occasion, be manipulated so that the operator can realize a cooperative concurrence benefit (Wickens and Boles, 1983).

To provide a framework for subsequent discussion, it will be useful to provide an explicit definition of the concept of a task. In this paper, we define tasks in terms of the assignments given the subject (e.g. track a target on a CRT, memorize a list of words, discriminate among several stimuli) and the dependent measures that are used to assess the performance of the assignments (e.g. RT, tracking error, percent correct). The assignment is made in terms of a set of stimuli, a set of responses, and a set of rules that map the response set to the stimulus set. The size of the stimulus and the response sets can vary with the task. Moreover, some stimuli in the set may be more critical to the assignment than other stimuli. The criticality of the stimuli to any task depends on the nature of the stimuli and on the rules which define the assignment. For example, a subject must attend to an unpredictable tracking target in order to minimize tracking error. It is critical for this discussion to add that while the

task designer can control the definition of the task, the exact stimuli which a subject chooses to process may vary as a function of the subject's strategy. The interaction of environmental stimuli and subjects processing strategies is illustrated by the study of the effects on human performance of redundant combinations of stimulus dimensions. For example, in the speeded classification paradigm subjects are instructed to sort decks of cards quickly and accurately into one or more mutually exclusive categories on the basis of a single dimension (Garner, 1969; 1970). In one experimental condition, a second dimension provides redundant information (e.g. color and brightness). Which dimension does the subject choose to process on a particular trial? We don't know, although we infer that both are used, because the redundant dimensions are sorted faster and more accurately than either dimension alone.

We consider subjects to be in a dual-task situation when two separate assignments are given, each with its associated performance criteria. However, the boundary between dual and single task situations is sometimes rather fuzzy, especially when one criterion must be met in meeting the other. For example, in flying an aircraft it is essential to communicate with the air traffic controller to receive directions for approach and landing. In this case it is dificult to determine where one task terminates and the other task begins since successful manuvering of the aircraft depends on the directional and sequencing information received from the controller.

In the present research, we are particularly interested in the situation in which there are two clearly defined and separate performance criteria, and therefore two distinct tasks. Under some dual-task conditions, the processing of one task may prove benefical to the processing of another

task. For example, subjects may be required to perform concurrently two separate tasks. One task requires tracking a target with a cursor along a single axis on a CRT while the other task calls for a discrimination between flashes that differ in brightness. What if an event embedded in one of the tasks predicts with some degree of certainty the appearance of a change in events associated with the other task? For example, the spatial position of the tracking target may predict (i.e. may be correlated with) the brightness of the secondary task stimulus. Thus, while the assignments to the subject when specified in terms of the proper responses to the two stimulus sets have not changed, an overlap has been introduced between the two stimulus sets. The correlation between the target position and flash brightness, incorporates the tracking stimuli into the brightness discrimination task. Such overlap between tasks may sometimes facilitate, and sometimes hinder, performance in a dual task situation. In this study we examine some of the conditions that allow operators to benefit from such concurrency (see Navon and Gopher, 1979).

Several factors have been proposed to influence the degree of integrality between tasks. One factor, the *redundancy between components of the tasks*, has been described above. The redundancy between stimulus dimensions has also proven useful to subjects in performing perceptual and cognitive tasks. Garner and associates, through a series of studies employing a diverse set of measurement techniques, have drawn a distinction between pairs of stimulus dimensions that are defined as integral and those that are separable (Garner, 1969, 1974; Garner and Flowers, 1969; Garner and Felfoldy, 1970). Two dimensions are said to be integral if there must be a level specified on one dimension in order for a level on the other dimension to be realized (Garner, 1970). For example, the dimensions of brightness and

saturation of a color chip are clearly integral since one dimension cannot be specified without the other. The processing of integral dimensions is enhanced when the dimensions are correlated. Orthogonal combinations of integral dimensions result in decrements in performance when one dimension is to be processed and the other one ignored. Performance with separable dimensions is not affected by the relations between the dimensions. Thus, integral dimensions appear to be processed as a unit and therefore benefit from redundant combinations while being relatively difficult to attend selectively. In the context of the present study it is of interest to examine the relationship between dimensional integrality and dual-task integrality. Does dual-task performance benefit from redundancy between components of the tasks regardless of the relationship between dimensions or do these two factors interact?

Recent models of attention have emphasized the influence of the spatial location of stimulus properties on the efficiency of processing. Treisman (1977), in her Feature Integration Model of Attention, has argued that features which occur within the same central fixation of attention are processed in parallel and combined to form a single object. Once the object has been formed it is perceived and stored in memory as such. The importance of the spatial location of stimulus properties has also been emphasized in the research on integral and separable dimensions. By definition, integral dimensions must occur in the same space and time (Lockhead, 1966). In the context of the present study, it is of interest to determine if the spatial proximity of different tasks influences dual-task performance, and, if so, how spatial proximity interacts with other factors that affect the processing of dual-tasks.

Investigators have reported that the processing of several stimulus

dimensions is more efficient if the dimensions are incorporated in a single object rather than in several objects. In one such study subjects were to detect a different dimension on each of three separate objects, the same dimensions on three separate objects or three different dimensions on the same object (Lappin, 1967). Identification of the dimensions was best when the three dimensions were located on the same object. Kahneman and Treisman (1984) have investigated further the advantages of incorporating different dimensions within a single object. In their study subjects were instructed to read a word which was either adjacent to, or was surrounded by, a rectangular frame. The subjects' primary task was to read aloud the tachistoscopically presented word as fast as possible. The secondary task required the detection of a gap in the rectangular frame. The condition in which the word was surrounded by the frame was assumed to produce a perceptual object. The reaction time for reading was 23 msec longer when the frame and word were separate than when the frame surrounded the word. Subjects were also significantly more accurate in locating the gap when the frame surrounded the word (84%) than when the two were separate (73%). In both studies, superior performance resulted from a change in the relations among the properties of the stimuli, though the type of processing required by the tasks remained the same. When the entities to be processed appeared as properties of a single object, or context, performance was enhanced.

Kahneman and co-workers (Kahneman & Henik, 1981; Kahneman & Treisman, 1984; Kahneman & Chajczk, 1983; Kahneman, Treisman & Burkell, 1983) have underscored the importance of the object in attention by suggesting that attentional competition arises between, not within object files. This argument implies that tasks which require the processing of different dimensions of the same object will be processed within the same resource

framework. Tasks which necessitate the processing of separate objects will compete for processing resources. Thus, the degree to which two separate tasks can be integrated into a single object will presumably determine the resource competition between the tasks. The hypothesis of competition for resources between objects and not among different dimensions of the same object will be investigated further in the present study by manipulating the stimulus relations among dual-task pairs. The P300 component of the event-related brain potential (ERP), that has been shown to be a sensitive index of processing resources (see below), will be employed as one measure of resource allocation.

A second assumption of Kahneman's object file model of attention is that the allocation of attention to any aspect of the object file facilitates the production of all responses associated with the separable properties of the object. This implies that the relevant as well as irrelevant dimensions of the object are processed. The processing of irrelevant aspects will be carried out regardless of their effect on the relevant task. Thus, in some cases the additional processing will have facilitating effects on performance while in other cases interference between the relevant and irrelevant aspects will be produced. Indeed, there is ample evidence for patterns of both facilitation and interference (Pomerantz and Garner, 1973; Reicher, 1977; Stroop, 1935; Weisstein and Harris, 1974). In these studies, the stimuli are generally presented briefly and at low levels of illumination. Other studies, which have presented suprathreshold stimuli at durations exceeding 200 msec, have found that subjects are capable of selectively attending to particular properties of objects (Donchin and Cohen, 1967; Kramer, Wickens and Donchin, 1983). In the present study we predict that selective attention can be directed to

specific dimensions of objects since the objects are readily perceptible and the subjects are not stressed for a speeded response.

Insert Figure 1 about here

The relations between performance and processing resources form the basis for the dual-task performance models described above. Therefore, it is useful to briefly describe the types of relations which have been proposed. Figure 1 illustrates the different relations among resource functions that are presumed to underly performance in different dual-task situations. The left side of the figure depicts the change in performance of a primary and a secondary task as primary task difficulty increases. Three cases are examined: resource reciprocity, separate resources and dual-task integrality. These performance functions correspond to the functions shown on the right side of the figure which map primary task difficulty to the hypothetical allocation of resources between the two tasks. It is assumed that primary task performance is not influenced by primary task difficulty (Rolfe, 1971; Wickens, 1979). Although stable primary task performance is the ideal, it is rarely obtained in practice, especially with manipulations of system parameters (Kramer et al., 1983). Thus, in reality it can be expected that primary task performance will decrease with increases in the difficulty of the primary task. In the resource reciprocity case, increasing the difficulty of the primary task results in a decrease in the performance on the secondary task. The corresponding resource-difficulty function displays a tradeoff between the two tasks. Increasing primary task difficulty leads to an increased demand for resources by the primary task. This results in a decreasing supply of resources available for the secondary

---

Insert Figure 9 about here

The ERPs elicited by the probes employed during session 1 served as a baseline for the secondary task probes used in the later experimental sessions. In the first session, subjects were instructed to ignore the probes and concentrate on performing the tracking task. Thus, the P300s elicited by the probes in session 1 provide an index of subjects' ability to ignore extraneous stimuli while performing a task. Figure 9 presents the average parietal ERPs elicited by the probes during the performance of the tracking task in the practice session. A comparison of the waveforms in figure 9 with those in figure 8 illustrates the relatively small size of the ERPs elicited by the uncounted probes. This is especially apparent in the epoch associated with the P300 component. The waveforms presented in figure 9 present no evidence of a P300. Furthermore, the waveforms elicited by the uncounted probes do not discriminate among levels of tracking difficulty in any of the experimental conditions.

This result confirms our prediction that the ignored stimulus properties will not elicit a P300. This effect is obtained regardless of the relationship of the probe stimuli to the primary task objects. Therefore, based on the P300 amplitude measure, it appears that subjects are capable of directing their attention to one property of an object while ignoring another property of the same object (for additional evidence see; Donchin and Cohen, 1967; Heffley, Wickens and Donchin, 1978; Kramer et al., 1983).

<u>Secondary Task Probes: Correlated Dual-Tasks</u> The analysis and discussion of the ERPs elicited by the secondary task probes has thus far

---

Insert Figure 8 about here

The opposite effect of system order on P300 amplitude was predicted for the cursor and horizontal bar conditions. This prediction derives from the resource structure inferred from the Object File Model of Attention (Kahneman & Henik, 1981). We hypothesized that if two tasks required that the subjects process different properties of the same object then the resource structure of the two tasks would be similar. The direct relationship between P300 amplitude and system order for the primary task events and cursor probes is consistent with this hypothesis. It was also argued that if two tasks required the processing of different objects and these tasks overlapped in their resource demands as defined by the Multiple Resource Model, then the relationship between P300 amplitude and system order would be reciprocal between primary and secondary tasks. This hypothesis was confirmed with the dual-task combination of the tracking task and horizontal bar. Thus, the results obtained in the present study are consistent with both hypotheses concerning the resource structure of dual-tasks. When two tasks require the processing of different properties of the same object then the amplitude of the P300s elicited by stimuli associated with each task will change in the same direction with changes in system order. If, on the other hand, the two tasks require the processing of different objects then as the amplitude associated with one increases, the amplitude associated with the other will decrease. That is, we will obtain a reciprocity in P300 amplitudes for concurrent tasks that require the processing of different objects.

noteworthy. In all of the experimental conditions the single task count block elicits a large positivity at approximately 400 msec post-stimulus. This positive deflection has been identified as the P300 component. The three levels of system order elicit varying degrees of positivity which appear to depend on the particular experimental condition. For example, for all experimental conditions in which the secondary task probe is the cursor, the waveforms are most positive for the second order condition, of intermediate amplitude in the first/second order condition and smallest in amplitude in the first order condition ($F(2,22)=28.1$, $p<.001$), a trend that mirrors the P300s elicited by the primary task probes as illustrated in Figure 6. This sequence of levels of system order does not appear to be influenced by the position of the secondary task probe relative to the tracking task or the type of discrimination required of the subject. In the two conditions in the lower left of figure 8 in which the horizontal bar is counted and is located below the tracking task the sequence of the ERPs elicited by different levels of system order is clear and consistent. However, the order is the inverse of that obtained in the cursor conditions. The first order tracking condition elicits the largest positivity, the first/second order condition elicits an intermediate level of positivity and the second order condition produces the smallest amplitude ($F(2,22)=24.2$, $p<.001$). This trend in P300 amplitude is typical of secondary task probe stimuli (Isreal et al., 1980; Natani and Gomer, 1981). Finally, in the two conditions in which the horizontal bar is superimposed on the tracking task the ERPs elicited by different levels of system order are not significantly different from each other.

of stimuli x 2 task positions x 2 secondary tasks x 3 levels of system order x 3 electrodes) each composed of 128 time points. Each of the ERPs represents an average of 50 to 60 single trials. The PCA was performed on the covariance matrix of time points. Figure 7 shows the Varimax rotated component loadings for the first three components extracted by the PCA. The three components accounted for 79 percent of the variance in the covariance matrix. The component scores computed from a linear combination of time points by loading coefficients were then subjected to a repeated measures ANOVA.

ERP components are customarily defined in terms of their latency relative to a stimulus or response, electrode distribution, and sensitivity to experimental manipulations. Component 3 becomes increasingly positive from Fz to Pz ($F(2,22)=115.08$, $p<.001$) and the component loadings were maximal in the epoch associated with P300 (450 - 700 msec). Based on these criteria component 3 can be identified as the P300 (Donchin, Kramer and Wickens, 1982). The amplitude of the P300 was influenced by the system order of the tracking task. Increases in system order produced increases in the amplitude of the P300 component ($F(2,22)=12.84$, $p<.001$). Thus, consistent with previous research, the amplitude of the P300s elicited by discrete changes in a primary task increase with increases in the difficulty of that task. None of the other main effects or interactions attained statistical significance. Since the P300 component of the ERP represents the major focus of the experimental hypotheses, other components will not be discussed in the present paper.

<u>Secondary Task Probes: Uncorrelated Dual-Tasks</u> Figure 8 presents the average parietal ERPs elicited by the secondary task probes during the performance of the step tracking task. Several aspects of the waveforms are

The accuracy with which subjects counted the secondary task probes was not significantly affected by any of the experimental manipulations. Subjects' counting accuracy exceeded 97 percent in all of the experimental conditions.

Event-Related Brain Potentials

The treatment of the ERP data is divided into two sections. The first section examines the ERPs elicited by changes in the spatial position of the tracking target. The second section is concerned with the effects of the experimental manipulations on the ERPs elicited by the secondary task probes in the correlated and uncorrelated dual-task conditions.

Primary Task Events Figure 6 presents the ERPs elicited by changes in the spatial position of the tracking target in the dual-task conditions for the parietal recording site. It is evident that the ERPs differ in the amplitude of the positive components as the difficulty of the primary task is varied. This amplitude difference appears as early as 350 msec after the stimulus and continues to the end of the recording epoch. Across all conditions, it appears that the largest positivity is elicited when tracking is the most difficult, a trend that replicates the basic finding of Wickens et al. (1983).

Insert Figure 6 and 7 about here

The ERPs acquired in the dual-task conditions were quantified by averaging the single trials within experimental conditions and analyzing these averages by a Principal Components Analysis (PCA) technique (see Coles, Gratton, Kramer and Miller, in press; Donchin & Heffley, 1979). The data base submitted to the PCA consisted of 864 ERPs (12 subjects x 2 types

were required to perform the secondary task by counting changes in the horizontal bar ($F(1,11)=7.1$, $p<.05$). The differential effect of the type of secondary task object on tracking performance may be due to the relationship of the objects to the primary task and is consistent with the task integration hypothesis as set forth in Figure 1. The cursor is clearly a necessary component of the tracking task while the horizontal bar is not necessary for primary task performance. Thus, subjects may find it more difficult to track and count probes if the probes are extraneous to the tracking task than if the probes occur within the primary task stimuli. If this interpretation is correct we would expect that integration of the two tasks, achieved by correlating events in the primary task with events in the secondary task, would reduce the differences in RMS error between the two conditions. A comparison of the correlated and uncorrelated dual-task pairs supports this interpretation. The difference in RMS error between the horizontal bar and cursor conditions was eliminated when the primary and secondary tasks were correlated.

Average ratings of difficulty for each level of system order in the dual-task conditions are presented in Figure 5b. Subjects' perception of difficulty increased from the single task count condition to the dual-task conditions as well as with increases in system order within the dual-task conditions ($F(3,33)=44.39$, $p<.001$). Subjects rated the difficulty of the dual-tasks higher when performing the secondary task with the horizontal bar than they did when counting the intensity or translational changes of the cursor ($F(1,11)=13.84$, $p<.01$). Subjective ratings of difficulty did not differ between objects in the correlated dual-task conditions. Thus, subjects ratings of tracking difficulty are consistent with their overt performance, as measured by RMS tracking error.

position and different object - same position conditions were highly correlated (.85). Each of the dual-task blocks lasted approximately 6 min. Subsequent to the dual-task blocks subjects again performed three single task tracking blocks. ERPs, subjective ratings and RMS tracking error were recorded during the experimental sessions. The order of the experimental blocks was counterbalanced across subjects.

RESULTS

Performance Measures and Subjective Ratings

Figure 5a presents the RMS tracking error for each level of system order during dual-task performance. The figure suggests that increasing system order results in increases in subjects' tracking error. Planned comparisons indicated that subjects performed significantly better with first order than they did with first/second order tracking ($F(1,11)=5.64$, $p<.05$). Performance was also better in the first/second order condition than it was during second order tracking ($F(1,11)=8.58$, $p<.05$). The effect of system order on RMS error did not differ significantly across dual-task, single task or correlated tracking conditions. Thus, the secondary task did not intrude on primary task performance.

Insert Figure 5 about here

Although the secondary task did not affect the relationship between RMS error and system order in the single and dual-task tracking blocks, the type of secondary task object did influence the subjects' tracking error as is illustrated in Figure 5a. Tracking error was significantly lower when the secondary task involved counting changes of the cursor than when subjects

and secondary task stimulus objects (same or different objects), the spatial position of the primary and secondary task displays (same or different) and the type of secondary task (intensity or translational discriminations). The degree of correlation between the primary and secondary tasks was also manipulated, although this manipulation was not orthogonal to the other four factors. Subjects performed the dual tasks with either low or high (0 or .85, respectively) correlation at each level of difficulty in the same object - same position and different object - same position conditions with the intensity discrimination secondary task.

Procedure

Each of the twelve subjects participated in all of the experimental conditions. One practice and two experimental sessions, run on successive days, were required to complete the experiment. The practice session included 24 blocks of tracking and six secondary task count blocks. Each of the tracking blocks lasted four min. Subjects performed eight blocks of tracking at each of the three levels of system order. Secondary task blocks lasted approximately six min. Although subjects did not count the probes in the tracking blocks, ERPs were digitized from both step changes of the target and presentations of the probes. Thus, these blocks served as practice as well as an indication of subjects allocation of processing resources between the tracking task and the irrelevant probe stimuli.

The experimental sessions began with three tracking blocks, each lasting approximately 3 min. Following the tracking blocks, subjects performed 15 dual-task blocks. The 30 dual-task blocks divided between sessions 2 and 3 consisted of 24 blocks from the (3 tracking difficulty levels x 2 types of stimuli x 2 task positions x 2 secondary tasks) factorial design and 6 blocks in which dual-tasks in the same object - same

In the dual-task blocks subjects performed both the tracking and the count tasks. At the conclusion of each block of trials subjects reported their total count. At this time subjects also rated the subjective difficulty of the block on a bipolar scale from 1 (easy) to 7 (difficult). Following each block the subjects were informed of their count accuracy and root mean square (RMS) tracking error.

Recording System

EEG was recorded from three midline sites (Fz, Cz and Pz) and referred to linked mastoids. Two ground electrodes were positioned on the left side of the forehead. Burden Ag-AgCl electrodes affixed with collodion were used for scalp and mastoid recording. Beckman Biopotential electrodes, affixed with adhesive collars, were placed below and supra-orbitally to the right eye to record electro-oculogram (EOG) and this type of electrode was also used for ground recording. Electrode impedances did not exceed 5 kohms/cm.

The EEG and EOG were amplified with Van Gogh model 50000 amplifiers (time constant 10 sec and upper half amplitude of 35 Hz, 3dB octave roll-off). Both EEG and EOG were sampled for 1280 msec, beginning 100 msec prior to stimulus onset. The data was digitized every 10 msec. ERP's were filtered off-line (-3dB at 6.29 Hz, 0dB at 14.29Hz) prior to statistical analysis. Evaluation of each EOG record for eye movements and blinks was conducted off-line. EOG contamination of EEG traces was compensated for through the use of an eye movement correction procedure (Gratton, Coles & Donchin, 1982).

Design

A repeated measures, four way factorial design, was employed. The factors were primary task difficulty (count only, first order, first/second order and second order control dynamics), the relationship between primary

manipulation: (1) in the relatively easy condition *a* was set to zero, a pure first order (velocity) system, (2) in the moderate difficulty condition *a* was set to .5, a 50/50 combination of first and second order dynamics, and, (3) in the difficult tracking condition *a* was set to 1.0, a pure second order (acceleration) system. Numerous investigators have validated the increasing resource demands of higher order control (Kramer et al., 1983; North, 1977; Trumbo, Noble and Swink, 1967; Vidulich and Wickens, 1981; Wickens, Derrick, Micallizi and Beringer, 1980).

Insert Figures 3 and 4 about here

The subjects secondary task involved counting the total number of occurrences of a relevant probe. Probes were presented in a Bernoulli series. The probability of either of the stimuli occuring on any one trial was .50. In different experimental blocks, subjects counted the bright flashes of a horizontal bar, bright flashes of a cursor, translational changes of the cursor upward or translational changes of a horizontal bar downward (see Figures 3 and 4). The two types of stimulus events (brightness and translational changes) were equated for difficulty prior to the experiment. Secondary task probes occurred either on the same horizontal axis as the tracking task or 2 cm (1.5 degrees of visual angle) below it. A probe was presented every 3.6 to 4 sec. The presentation of the probe was temporally constrained so that it occured 1.8 to 2 sec subsequent to a step change in the tracking target. Thus the temporal sequence of the presentation of the probes (secondary task stimuli) and changes in the spatial position of the tracking target was fixed, while the temporal interval between these stimuli was variable.

METHOD

Subjects

Twelve right handed persons (6 male and 6 female) were recruited from the student population at the University of Illinois and paid for their participation in the study. None of the students had any prior experience with the pursuit step tracking task. All of the subjects had normal or corrected to normal vision.

Insert Figure 2 about here

Step Tracking and Discrimination Tasks

The single axis pursuit step tracking task is illustrated in Figure 2. The tracking display which consisted of the computer driven target and the subject controlled cursor was presented on a Hewlett Packard CRT which was positioned approximately 70 cm from the subjects. The target and cursor were 1.2 cm x 1.2 cm in size and subtended a visual angle of 1.0 degrees. The target changed its position along the horizontal axis once every 3.6 to 4 sec and the subjects' task was to nullify the position error between the target and cursor. The cursor was controlled by manipulating a joystick with the right hand. Pursuit step tracking was defined as the primary task. The dynamics for the tracking stick were composed of a linear combination of first order (velocity) and second order (acceleration) components. That is, the system output, X(t), is represented by the following equation.

$$X(t) = [(1-a)\int u(t)\,dt] + [(a)\iint u(t)\,dt]$$

where: u = stick position; t = time and a = difficulty level.

The task was conducted at three different levels of the system order

presumed to consume resources which would have normally been used in the processing of the secondary task. Thus, the secondary task P300's mirror the proposed resource function. If P300 does in fact reflect the resource structure of dual-tasks then it would be predicted that P300s elicited by primary task events would increase in amplitude with increases in the difficulty of the primary task. This hypothesis was confirmed in a study in which P300s were elicited by discrete spatial changes in the position of a target in a tracking task (Wickens, Kramer, Vanasse & Donchin, 1983). Increasing the difficulty of the tracking task by decreasing the stability of the control dynamics resulted in a systematic increase in P300 amplitude, a finding similar to that depicted in the top right of figure 1.

The reciprocal relationship between P300s elicited by primary and secondary task stimuli as a function of primary task difficulty is identical to the resource tradeoffs presumed to underly dual-task performance decrements. Thus, the hypothetical resource functions illustrated on the right side of figure 1 might be inferred from changes in the amplitude of P300 as a function of task difficulty. In the present experiment P300s will be employed to provide information concerning the resource framework of dual-task combinations, as various perceptual characteristics are manipulated. Specifically we will manipulate the spatial proximity of the primary and secondary task displays, the degree to which the primary and secondary task displays constitute a common object, the similarity of resource demands of primary and secondary tasks, and the degree of correlation between primary and secondary task events. The first three factors will be varied in a factorial design, while the correlation manipulation will be varied at a fixed level of position and stimulus resource demand variables.

responses. In this study we used a psychophysiological index of resource allocation, the amplitude of the P300 component of the ERP.

The ERP is a transient series of voltage oscillations in the brain that can be recorded on the scalp in response to a discrete stimulus event (Donchin, 1975; Regan, 1972). The ERP has traditionally been partitioned into a number of separate components. In most cases component labels indicate both the polarity and approximate latency of the peak (e.g. N100 is a negative peak occurring approximately 100 msec after stimulus onset). Other relatively slow components, such as the slow wave (SW) and contingent negative variation (CNV), are labeled on the basis of their duration or relationship to the experimental arrangement. The amplitude and latency of the early components, those occuring within the first 100 msec, have been shown to be influenced by the physical attributes of stimuli (e.g. intensity, modality, presentation rate). These components have been labeled exogenous. Later components such as N200 and P300 are nonobligatory responses to stimuli. These endogenous components reflect the strategies, expectancies and other psychological processes of the subjects and are uninfluenced by the physical attributes of the stimuli. One such endogenous component, the P300, represents one of the dependent variables in the present study.

The P300 component of the ERP has been found useful in providing information concerning the allocation of resources to concurrently performed tasks. P300's elicited by discrete secondary task events decrease in amplitude with increases in the difficulty of the primary task (Isreal et al., 1980; Kramer, Wickens & Donchin, 1983). The secondary task methodology assumes that changes in primary task difficulty will be reflected in secondary task performance. Increasing the difficulty of a primary task is

task. In the separate resource case, increasing the difficulty of the primary task fails to affect performance on the secondary task. The corresponding resource-difficulty function reflects the insensitivity of the secondary task to the withdrawal of resources. In the separate resource case, the resources required for the performance of the primary and secondary tasks are not the same. In the dual-task integrality example, secondary task performance increases as a function of increasing primary task difficulty. Thus, it is assumed that the secondary task can benefit from the additional resources allocated to the primary task. The corresponding resource-difficulty function displays a single function which represents the resources allocated to both tasks.

Dual-task integrality has been described on two levels. On a performance level, dual-task integrality results in a facilitation in the performance of one or both tasks when executed concurrently. Facilitation is relative to conditions in which the two tasks are performed separately or when the stimulus relations but not the processing requirements change between dual-task pairs. On a resource level, dual-task integrality occurs when two tasks can be processed within the same resource framework. Thus, there appear to be at least two different types of dual-task combinations that do not result in performance tradeoffs. As argued by capacity theories, tasks which require different processing resources can be sucessfully time shared. In the present study we are suggesting that dual-task decrements can also be avoided if the two tasks permit integrated processing even if the tasks require the same type of processing resources.

ERPs and Processing Resources

One difficulty in resolving issues regarding dual-task integrality is a way of assessing resource allocation that is independent of the criterion

been concerned with dual-tasks which are uncorrelated. Can we expect the relationships observed with uncorrelated dual-tasks to generalize to situations in which the events in one task predict the events in the other task with some degree of certainty? What effects will inter-task correlation have on the resource structure of the two tasks? These questions are examined in the present section by analyzing the effects of inter-task correlation on the relationship between P300 amplitude and system order.

Figure 10 presents the average parietal ERPs elicited by the correlated and uncorrelated dual-task conditions when the cursor and bar are flashed. There are several interesting aspects of these waveforms. A comparison of the ERPs elicited in the correlated and uncorrelated cursor probe conditions suggests that system order has the same effect on the ERPs in both conditions. The ERPs elicited by the cursor probes during second order tracking possess a large positive amplitude, the P300. The first/second order condition waveforms are of intermediate amplitude and the first order condition ERPs are smallest in amplitude ($F(2,22)=10.9$, $p<.001$). An examination of the waveforms elicited by the horizontal bar probes presents a different picture. As noted previously, the effect of system order is not significant in the uncorrelated horizontal bar condition. However, the ERPs elicited in the correlated horizontal bar condition increase in positivity with increases in system order ($F(2,22)=12.3$, $p<.001$). Thus, it appears that the effect of system order on the ERPs is the same across the two cursor conditions and the correlated horizontal bar condition. Correlating the tracking and probe events performs the same "integrating" function on the processing as is accomplished by combining them into a common object.

Insert Figure 10 about here

These results suggest that when two tasks are already being processed within the same resource framework, as was the case for the uncorrelated dual-task cursor condition, correlation does not have a large effect on the resources allocated to the tasks. The relationship between P300 amplitude and system order was not significantly different in the correlated and uncorrelated dual-task conditions. Thus, when the two tasks require the processing of different properties on the same object, the processing of the tasks is in some sense integrated and inter-task correlation does not enhance this integrality further. However, when two concurrently performed tasks require the processing of separate objects, as was the case in the horizontal bar conditions, the presence of inter-task correlation does appear to enhance the integrality between tasks. This increase in dual-task integrality is inferred from the change in the relationship between P300 amplitude and system order in the correlated and uncorrelated horizontal bar conditions. P300 amplitude changes with system order in the correlated condition in the same manner that it does when P300 is elicited by primary task events, suggesting an overlap in the resource structure between tasks.

GENERAL DISCUSSION

In most dual-task combinations, increasing the difficulty of one task is assumed to consume resources that normally would have been employed in the processing of the other task. The resources shared by these two tasks are presumed to be reciprocal in nature. Thus, increasing the difficulty of one task leads to a decrement in performance on one or both of the

concurrently performed tasks. In other cases, the two tasks require different processing resources and therefore do not result in resource tradeoffs. These tasks are generally performed as well together as they are alone (Navon and Gopher, 1979; Wickens, 1984). Under conditions of dual-task integrality, the secondary task increases processing demands within the domain of the primary task. Therefore, in the case of dual-task integrality, resource reciprocity is not obtained although both tasks require the same resources. Dual-task integrality results in a facilitation in performance in one or both tasks when performed concurrently. Facilitation is relative to conditions in which the two tasks are performed separately or when the stimulus relations but not the processing requirements change between dual-task pairs.

Insert Figure 11 about here

The concept of dual-task integrality is operationally defined in the current context as occurring when the amplitude of the P300s elicited by secondary task probes increase with increases in the difficulty of the primary task. Four experimental variables were manipulated, each intended to foster increasing degrees of integrality between the primary and secondary tasks. The dataproved to be systematic and it is possible to order the variables in terms of the degree to which they fostered integrality. In discussing the data, reference is made to figure 11, in which P300 amplitude in each condition is shown as a function of the system order of the tracking task.

First and most consistent are the effects of the object properties on dual-task integrality as inferred from changes in the amplitude of P300.

When the relevant stimuli from the two tasks were part of the primary task object, integrality was observed at its maximum value. P300s elicited by secondary task events increased in amplitude with increases in the difficulty of the primary task. Given that the two tasks required the processing of different properties of the same object, neither a change in the specific properties (spatial or intensity) nor a change in the correlation between tasks could alter the degree of integrality. Furthermore, the object-derived benefit was also reflected by the RMS error data. Tracking performance was superior when the two tasks required the processing of different properties of a single object as compared to the processing of separate objects. These results are consistent with previous findings which suggest that different properties of an object tend to be processed in parallel (Kahneman and Henik, 1981; Lappin, 1967; Treisman, 1977). The important knowledge added by the present study is the direct measure of resource investment, and the characteristic that reciprocity is defined here in terms of a resource-demand manipulation and not just an absolute performance level.

Second, and equally strong, is the effect of correlation on dual-task integrality. When the two tasks are correlated, integrality is shown. P300s elicited by secondary task events which are correlated with events in the primary task increase in amplitude with increases in the difficulty of the primary task. When events in the two tasks are not correlated and the tasks require the processing of different objects, integrality is lost and reciprocity is sometimes shown. There are several reasons why correlation may produce integrality. Again, the object file concept may underlie this effect. Different properties of a single object are typically correlated as we experience them in the real world. So, turning this around, the

correlation of stimuli may foster object file perception and hence, dual-task integrality. Garner and co-workers have found the processing of integral stimulus dimensions is enhanced when the dimensions are correlated, purportedly because the integral dimensions function as a single unit (Garner, 1969; Garner and Felfoldy, 1970). In the present study it appears that two tasks which are correlated also seem to function as a "unit" and therefore benefit from the redundancy.

Thirdly, spatial location fosters integrality, although to a lesser extent than the properties of an object or the correlation between tasks. Of course if the two tasks require the processing of different properties of a single object this guarantees a common spatial location. However, even when there were different primary and secondary task objects (horizontal bar probes), we found that locating them together in space, while not producing integrality still reduced the level of reciprocity so that the P300 function was flat. Again, returning to real world experience, it is true that the properties of an object are typically close together in space; but proximity does not guarantee integrality. The ease with which subjects can focus on some information at a location in space while completely ignoring other information at the same location has been demonstrated in several experiments (Donchin and Cohen, 1967; Fischer, Haines and Price, 1980; Neisser and Beckman, 1975).

Finally, the one variable which did not produce integrality was the nature of the resources demanded by the secondary task probes. Whether the probes used common (spatial) or dissimilar (intensity) resources to the spatial processing underlying the primary task had no effect on the degree of integrality. Perhaps the processing of changes in intensity does not require a different type of resources from the processing of spatial

changes. In the case of both variables, subjects were required to detect changes in magnitude. It may be that the processing of these magnitude changes is accomplished through common spatial-analog resources. In fact, if the two variables had demanded separate resources then we should have seen P300 in the separate object condition to be more affected by increases in tracking difficulty when spatial rather than intensity probes were used. The fact that we did not, supports the argument that the processing of changes in intensity and spatial position can be accomplished with similar resources.

Thus far we have argued that when two tasks require the processing of different properties of a single object, integrality is observed. Other investigators have also found that it is difficult to selectively attend to one property of an object while ignoring other properties (Kahneman and Chajczyk, 1983; Stroop, 1935). This seems to be especially true if the two properties are integral, in the sense that for one property to be realized there must be a level specified on the other property (Garner, 1970). In the first session of the present study subjects were instructed to perform the tracking task and ignore the extraneous probes. The probes were changes in properties of the primary task objects that were not necessary for tracking performance. These probes became the secondary task stimuli in the later, experimental sessions. ERPs elicited by the ignored probes did not possess a P300 component. However, P300s were elicited by the probes when they represented a secondary task. Although the presence or absence of the P300 does not in and of itself indicate the success or failure of selective attention it does provide information concerning the amount of task related processing (Hillyard and Kutas, 1983; Picton et al., 1978). The P300 results suggest that the task relevant properties of the primary task objects were

processed to a greater extent than the irrelevant properties. Thus, it appears that when the properties of an object are highly discriminable, subjects are capable of selectively attending to specific properties of an object while ignoring others.

Insert Figure 12 about here

Figure 12 presents a model of the processing framework underlying the phenomenon of dual-task integrality as inferred from measures of P300 amplitude. Each of the three stimuli, the target, cursor and horizontal bar possess a number of properties. The subjects are instructed that some of the properties are task relevant and require processing while other properties are not necessary for successful performance of the tasks. The relevant properties are assigned a high processing priority while other properties receive a lower priority. Large P300s are elicited by the properties which are assigned a high priority, small P300s are elicited by the low priority properties. At a higher level of analysis the stimulus properties are then aggregated on the basis of task assignments and priorities. The properties that are necessary for primary task performance receive a higher processing priority than the properties for the secondary task. However, secondary task properties which occur on primary task objects are assigned the same processing priority as primary task properties. Thus, the processing of the secondary task properties is done within the domain of the primary task. As the primary task demands more resources, secondary task processing will benefit to the extent that it shares primary task properties. This process represents the phenomenon of dual-task integrality and is revealed by examining the graded effect of task difficulty on the amplitude of the P300.

Secondary task properties which do not occur on primary task objects are assigned a lower priority. These properties receive the resources remaining after primary task processing. This process is referred to as resource reciprocity. Resource reciprocity also depends on the overlap between the resources required for primary task performance and those needed for the performance of the secondary task. If the two tasks require different types of processing resources, resource reciprocity will not occur (Navon and Gopher, 1979; Wickens, 1980). Inter-task correlation and spatial overlap of the task relevant properties increase the integrality between tasks by decreasing the distance between the primary and secondary tasks on the integrality continuum. Inter-task correlation is more influential in this respect than physical proximity.

The changes in P300 amplitude as a function of *primary task difficulty* have been used to support a resource model of dual-task integrality. It may be argued that these results can be interpreted more parsimonously by a model of eye fixations. According to this argument, the structure of the tracking task encourages subjects to change their eye fixation strategy at different levels of system order and that changes in P300 amplitude reflect the length or frequency of fixation. It might be assumed that longer or more frequent fixations produce large P300s while short or infrequent fixations result in small P300s. Subjects may spend most of their time fixating on the target during first order tracking since the cursor is relatively easy to control in this condition. This strategy would predict small P300s to both the cursor and horizontal bar in the first order condition. In the second order condition in which the cursor is relatively difficult to control, subjects may spend the majority of their time fixating on the cursor. The target would consume most of the remaining fixation time. This second order

fixation strategy would predict that the cursor probe would elicit large amplitude P300s while the horizontal bars would result in even smaller P300s than in the first order condition. The results obtained in the present study are consistent with these predictions. If a simple eye fixation interpretation predicts the results then why bother postulating a more complex information processing model?

There are at least two arguments that question the adequacy of the fixation interpretation. First, in the condition in which the horizontal bar is superimposed on the tracking task the eye fixation interpretation would predict large P300s for the horizontal bars. The results obtained in the experiment disagree with this prediction. P300s were of intermediate size as compared to the other experimental conditions (see figure 11). The second argument against the eye fixation interpretation is based on the results of a subsequent investigation (Kramer, in preparation). In that study secondary task P300s were elicited by intensity changes in both the target and the cursor. The fixation model predicts that the relative size of the target and cursor P300s should vary as a function of system order; the target eliciting the larger P300 during first order tracking and the cursor eliciting the larger amplitude P300 in the second order condition. The results did not support the prediction. P300s for both the target and cursor increased in amplitude with increases in system order. This result is predicted by the resource model of dual-task integrality. Thus, it appears that the fixation interpretation does not provide a reliable account of the P300 results.

The resource framework inferred from the P300 provides a theoretical account of the effect of several factors on the phenomenon of dual-task integrality. The results also have practical implications. The P300 component has been employed as a measure of cognitive workload. P300s

elicited by secondary task stimuli decrease in amplitude with increases in primary task difficulty. P300s elicited by discrete primary task events increase in amplitude with increases in the difficulty of the primary task. The resources allocated to tasks have been inferred from changes in P300 amplitude. The results obtained in the present study suggest that the reciprocal relationship between the primary and secondary task depends on the structure of the dual-task. For example, the relationship between P300 amplitude and task difficulty changes from the case in which the two tasks require the processing of different properties of the same object to the situation in which the two tasks necessitate the processing of different objects. Furthermore, inter-task correlation and the physical proximity of task displays also have a significant effect on the resource structure of the dual-task pair. These findings suggest that a reliable analysis of the processing demands of a task can only take place within a theoretical framework. The model of dual-task integrality offers one such framework.

References

Alwitt, L.F. (1981). Two neural mechanisms related to modes of selective attention. Journal of Experimental Psychology: Human Perception and Performance, 7, 324-332.

Coles, M.G.H., Gratton, G., Kramer, A. and Miller, G. (in press). Principles of signal acquisition and analysis. In M.G.H. Coles, S.W. Porges and E. Donchin (Eds.), Psychophysiology: Systems, Processes and Applications. New York: Guilford Press.

Donchin, E. (1975). Brain electrical correlates of pattern recognition. In G.F. Inbar (Ed.), Signal Analysis and Pattern Recognition in Biomedical Engineering. New York: John Wiley.

Donchin, E. and Cohen, L. (1967). Average evoked potentials and intramodality selective attention. Electroencephalography and Clinical Neurophysiology, 22, 537-546.

Donchin, E. and Heffley, E. (1979). Multivariate analysis of event related potential data: A tutorial. In D. Otto (Ed.), Multidisciplinary Perspectives in Event-Related Potential Research(pp. 555-572). EPA 600/9-77-043, Washington, D.C.: U.S. Government Printing Office.

Donchin, E., Kramer, A. and Wickens, C. (1982). Probing the cognitive infrastructure with event-related brain potentials. In M.L Frazier and R.C. Crombie (Eds.), Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics. AFFTC-TR-82-5. Edwards Air Force Base, California.

Fischer, E., Haines, R. and Price, T. (1980). Cognitive issues in head-up displays. NASA Technical Paper 1711. Washington, D.C.: NASA.

Garner, W.R. (1969). Speed of discrimination with redundant stimulus attributes. Perception and Psychophysics, 6, 221-224.

Garner, W.R. (1970). The stimulus in information processing. American Psychologist, 25, 350-358.

Garner, W.R. (1974). The Processing of Information and Structure. Potomac, Maryland: Erlbaum Associates.

Garner, W.R. and Flowers, J.H. (1969). The effect of redundant stimulus elements on visual discriminations as a function of element heterogenity, equal discriminability and position uncertainity. Perception and Psychophysics, 6, 216-220.

Garner, W.R. and Felfoldy, G.L. (1970). Integrality of stimulus dimensions in various types of information processing tasks. Cognitive Psychology, 1, 225-241.

Gratton, G., Coles, M.G.H. and Donchin, E. (1982). A new method for off-line removal of ocular artifact. Electroencephalography and Clinical Neurophysiology, 55, 468-484.

Heffley, E., Wickens, C.D. and Donchin, E. (1978). Intramodality selective attention and P300: A reexamination in visual monitoring task. Psychophysiology, 15, 269-270.

Hillyard, S.A. and Kutas, M. (1983). Electrophysiology of cognitive processing. Annual Review of Psychology, 34, 33-61.

Isreal, J.B. (1980). Structural interference in dual-task performance: Behavioral and electrophysiological data. Unpublished Ph.D. Dissertation, University of Illinois.

Isreal, J.B., Wickens, C.D., Chesney, G.L. and Donchin, E. (1980). The event-related brain potential as an index of display monitoring workload. Human Factors, 22,

211-224.

James, W. (1890). Principles of Psychology Vol I. New York, Henry Holt and Company.

Kahneman, D. and Henik, A. (1981). Perceptual organization and attention. In M. Kubovy and J.R. Pomerantz (Eds.), Perceptual Organization(pp. 181-209). Hillside, New Jersey: Erlbaum.

Kahneman, D. and Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, and R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Kahneman, D. and Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of stroops effects by color irrelevant stimuli. Journal of Experimental Psychology: Human Perception and Performance, 9, 497-509.

Kahneman, D., Treisman, A. and Burkell, J. (1983). The cost of visual filtering: A new interference effect. Journal of Experimental Psychology: Human Perception and Performance, 9, 510-522.

Kramer, A., Wickens, C.D. and Donchin, E. (1983). An analysis of the processing demands of a complex perceptual-motor task. Human Factors, 25, 597-622.

Lappin, J.S. (1967). Attention in the identification of stimuli in complex visual displays. Journal of Experimental Psychology, 75, 321-328.

Lockhead, G.H. (1966). Effects of dimensional redundancy on visual discrimination. Journal of Experimental Psychology, 72, 95-104.

Natani, K. and Gomer, F.E. (1981). Electrocortical activity and operator

workload: A comparison of changes in the electroencephalogram and in event-related potentials. McDonnell Douglas Technical Report, MDC E2427. McDonnell Douglas Astronautics Company, St Louis: Missouri.

Navon, D. and Gopher, D. (1979). On the economy of the human processing system. Psychological Review, 86, 214-255.

Neisser, U. and Beckman, R. (1975). Selective looking: Attention to visually specified events. Cognitive Psychology, 7, 480-494.

North, R.A. (1977). Task components and demands as factors in dual-task performance. Aviation Research Laboratory. Report Number ARL-77-2/AFOSE-77-2. University of Illinois at Urbana-Champaign,

Picton, T.W., Campbell, K.B., Baribeau-Braun, J. and Proulx, G.B. (1978). The neurophysiology of human attention: A tutorial review. In J. Requin (Ed.), Attention and Performance VII. Hillsdale, New Jersey: Erlbaum Associates.

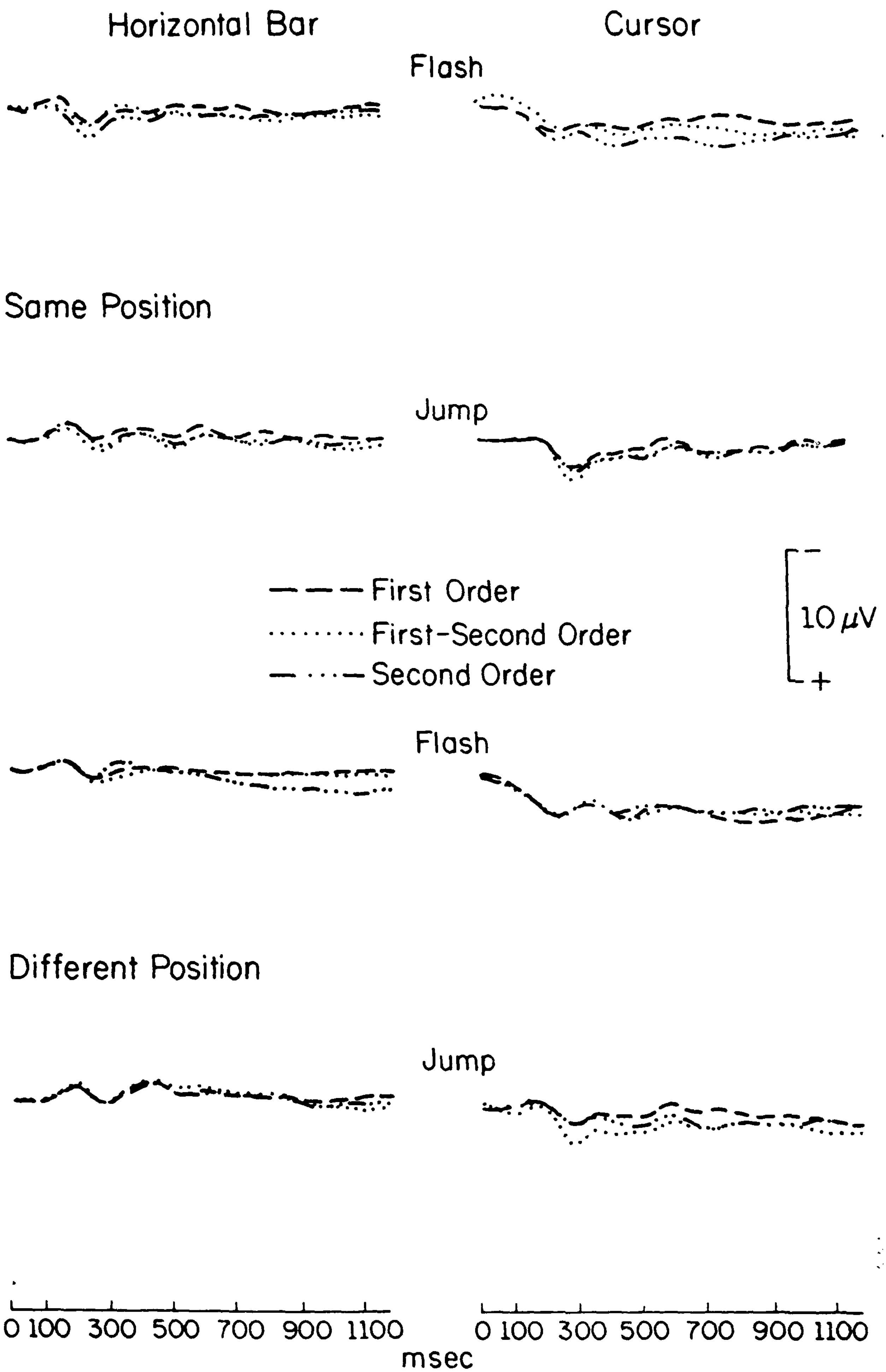
Pomerantz, J.R. and Garner, W.R. (1973). Stimulus configuration in selective attention tasks. Perception and Psychophysics, 18, 355-361.

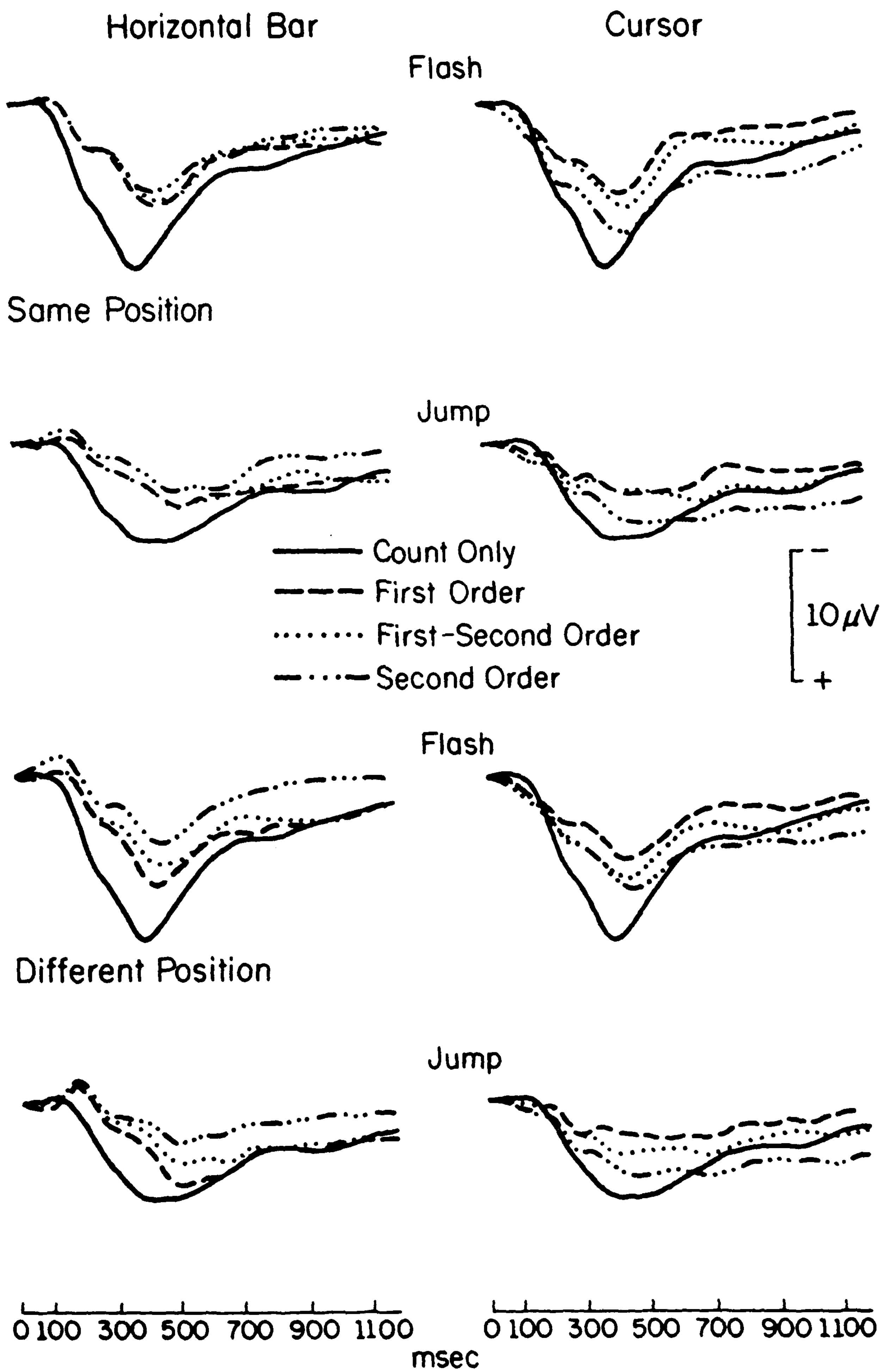
Regan, D. (1972). Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. London: Chapman and Hall.

Reicher, G.M. (1977). Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 81, 275-280.

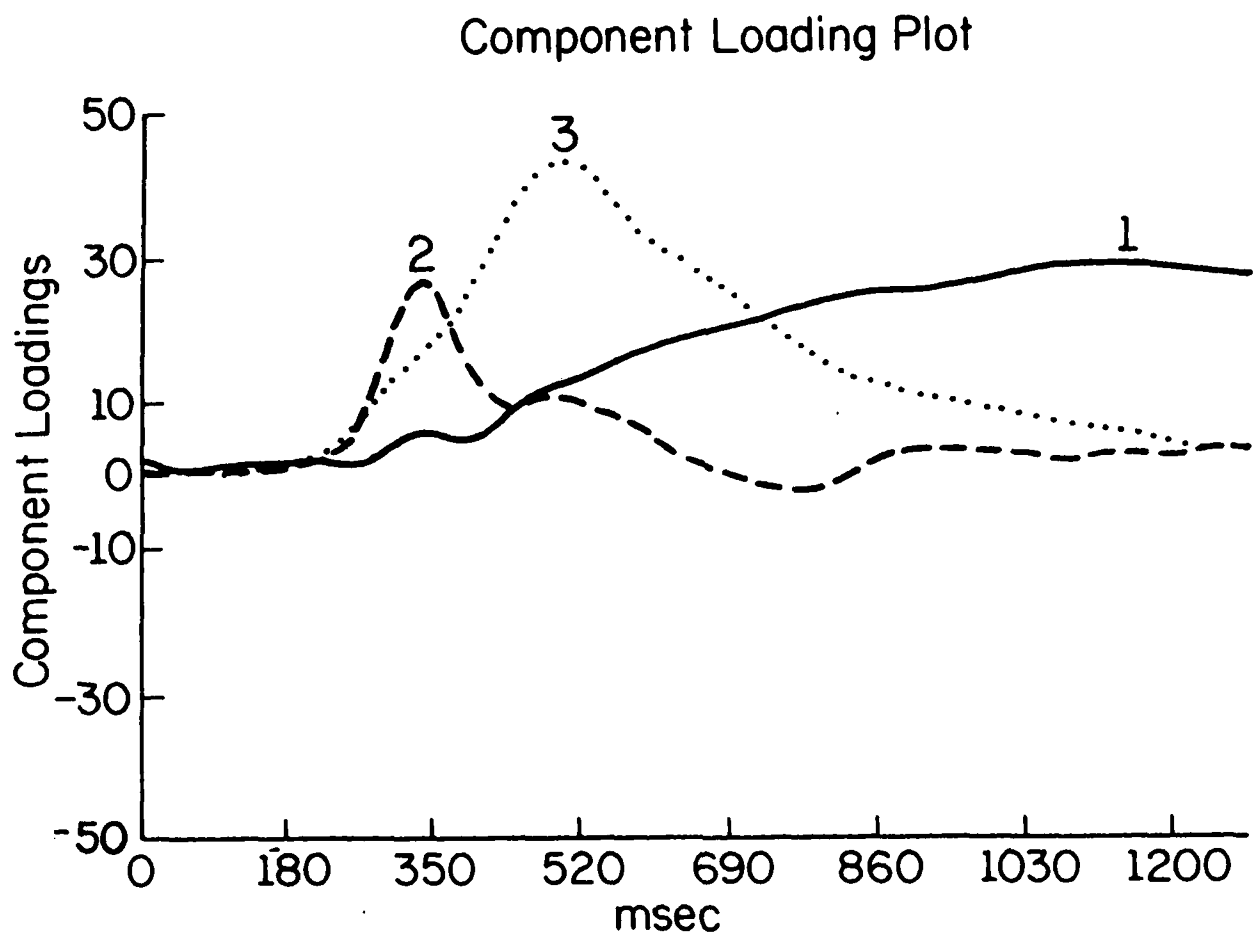
Rolfe, J.M. (1971). The secondary task as a measure of mental load. In W.T. Singleton, J.G. Fox and P. Witfield (Eds.), Measurement of Man at Work. London: Taylor and Francis.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions.

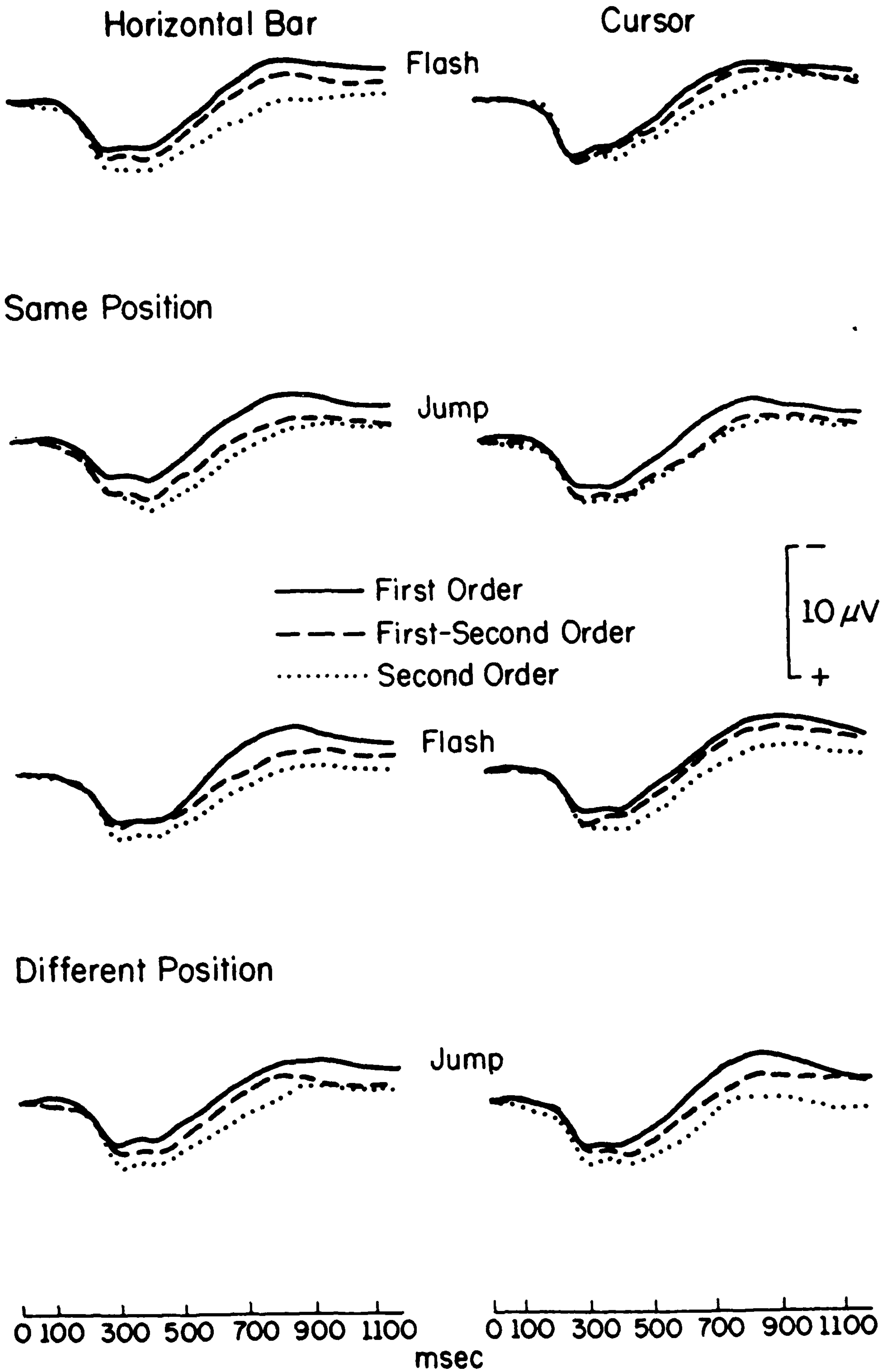

Horizontal Bar
Cursor
Flash
Same Position
Jump
First Order
First-Second Order
Second Order
10 μV
−
+
Flash
Different Position
Jump
0 100 300 500 700 900 1100
0 100 300 500 700 900 1100
msec

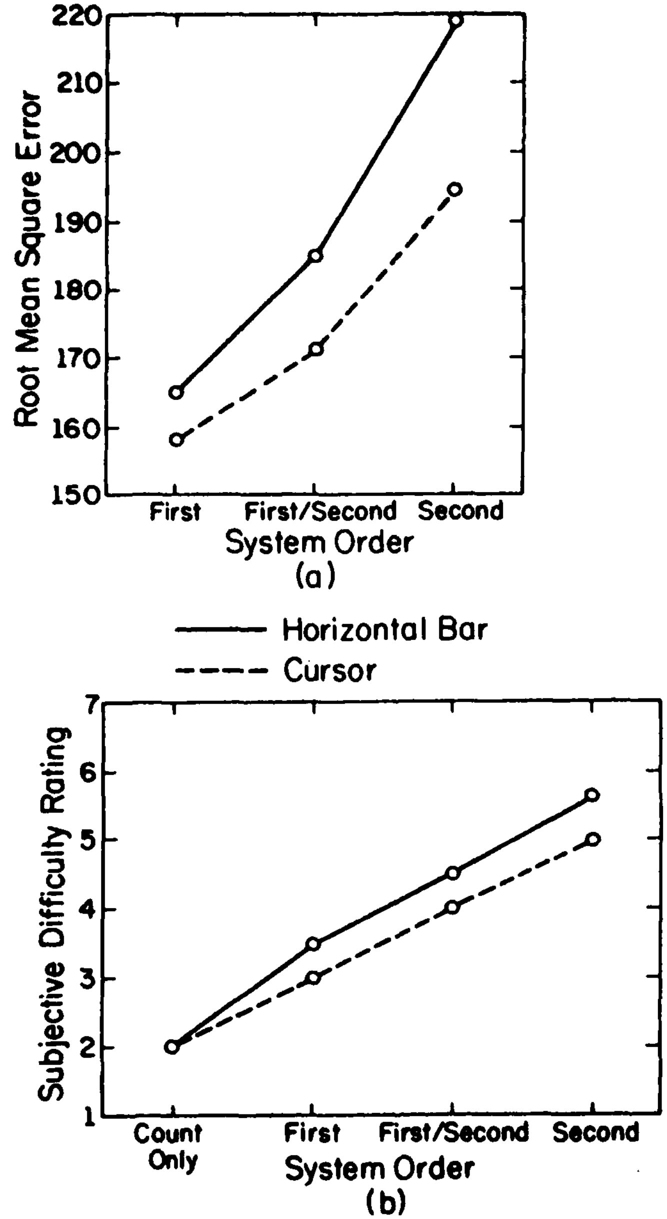


Horizontal Bar
Cursor
Flash
Same Position
Jump
Count Only
First Order
First-Second Order
Second Order
10 μV
−
+
Flash
Different Position
Jump
0 100 300 500 700 900 1100
0 100 300 500 700 900 1100
msec

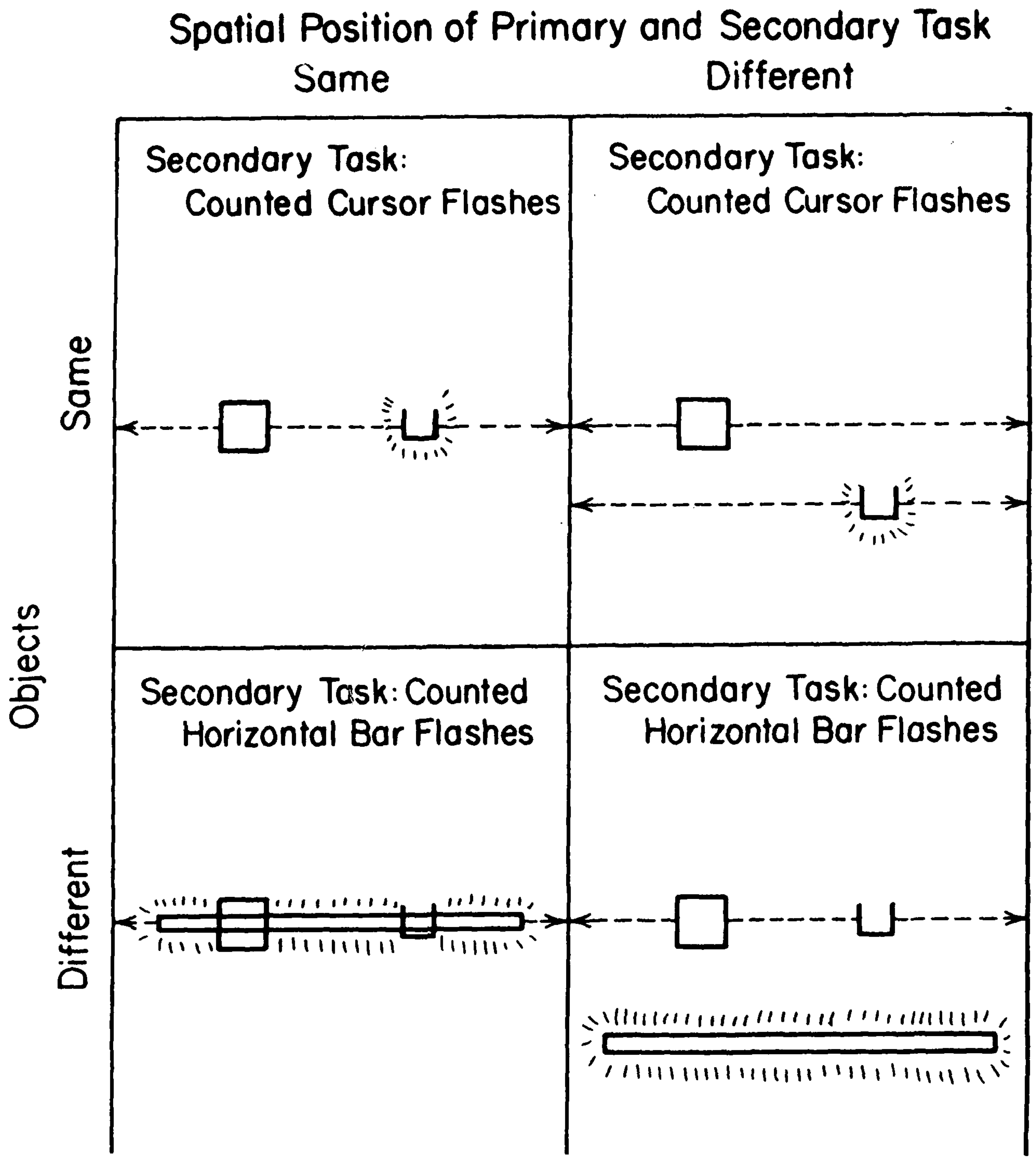


Component Loading Plot
Component Loadings
50
30
10
0
-10
-30
-50
0
180
350
520
690
860
1030
1200
msec
1
2
3

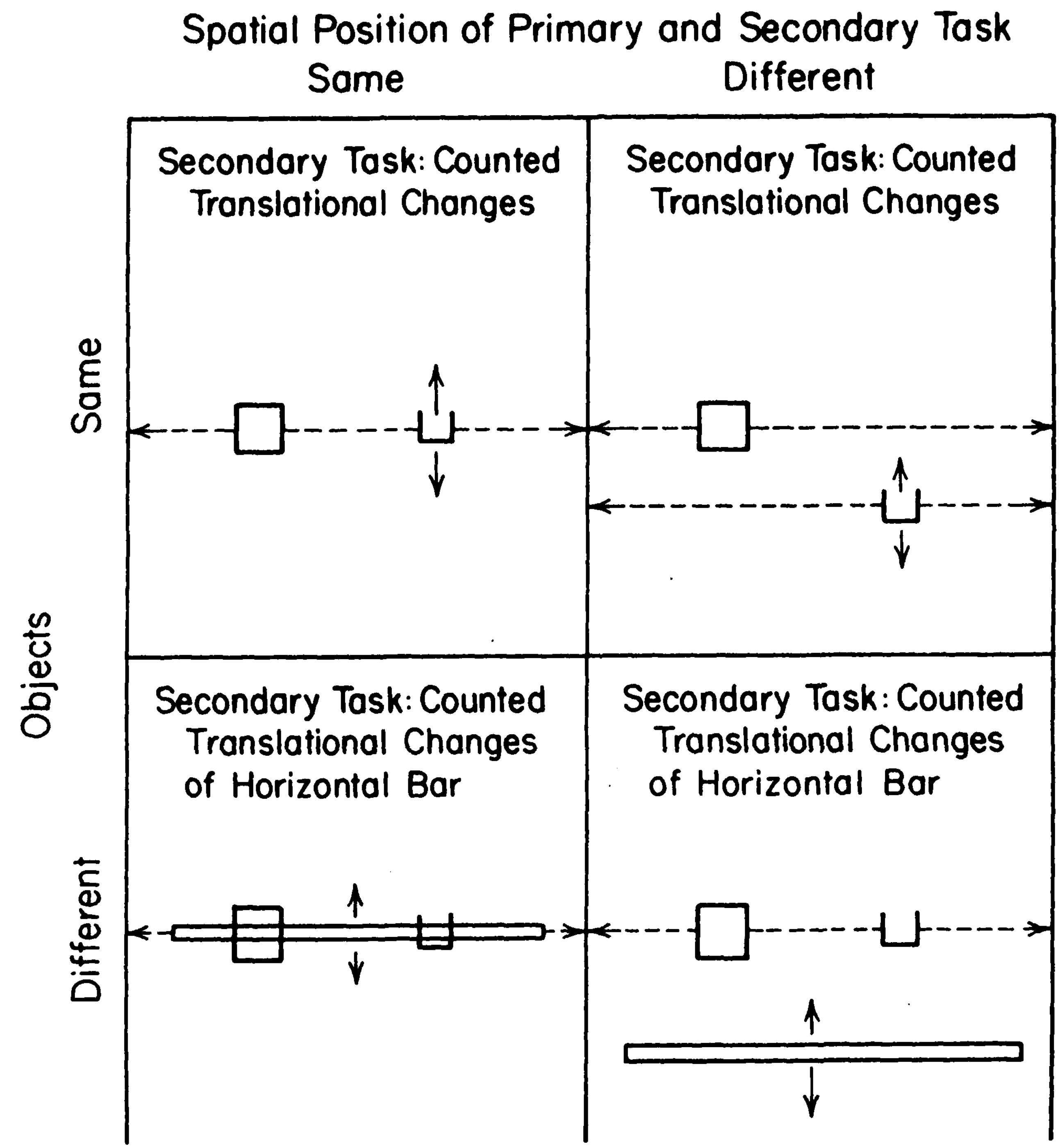

Horizontal Bar
Cursor
Flash
Same Position
Jump
First Order
First-Second Order
Second Order
10 μV
−
+
Flash
Different Position
Jump
0 100 300 500 700 900 1100
0 100 300 500 700 900 1100
msec


Root Mean Square Error
220
210
200
190
180
170
160
150
First
First/Second
Second
System Order
(a)
Horizontal Bar
Cursor
Subjective Difficulty Rating
7
6
5
4
3
2
1
Count Only
First
First/Second
Second
System Order
(b)

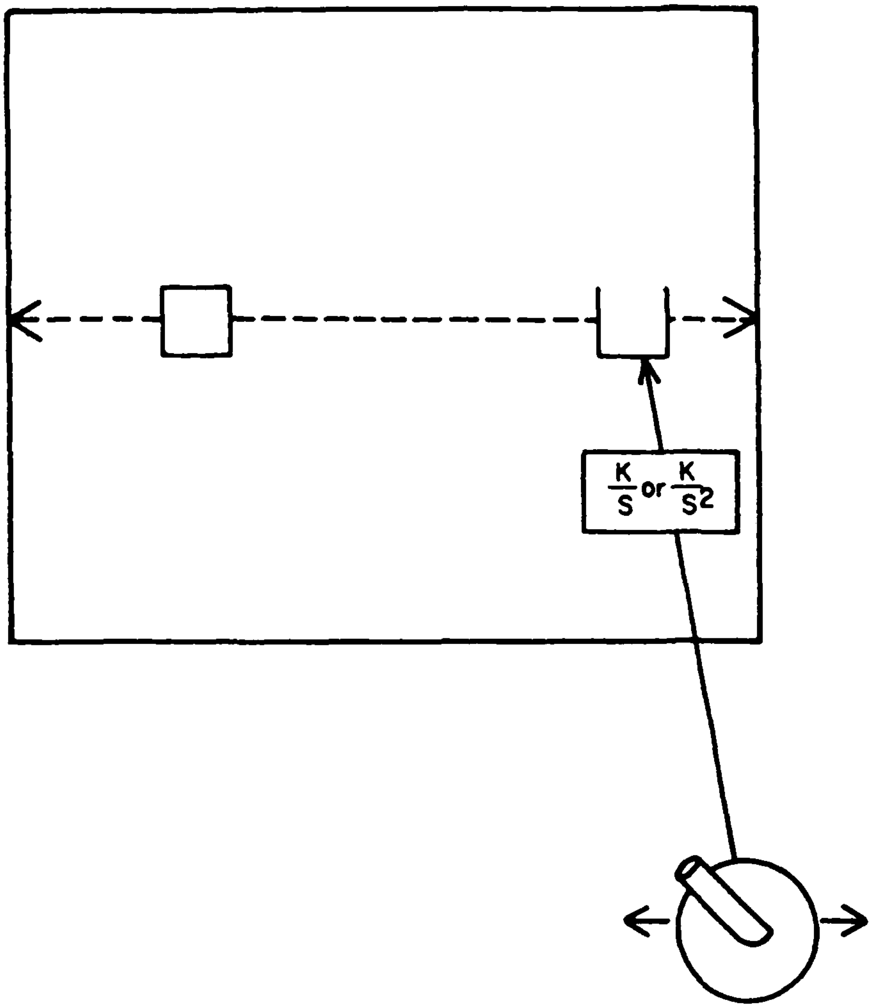
Intensity Discrimination

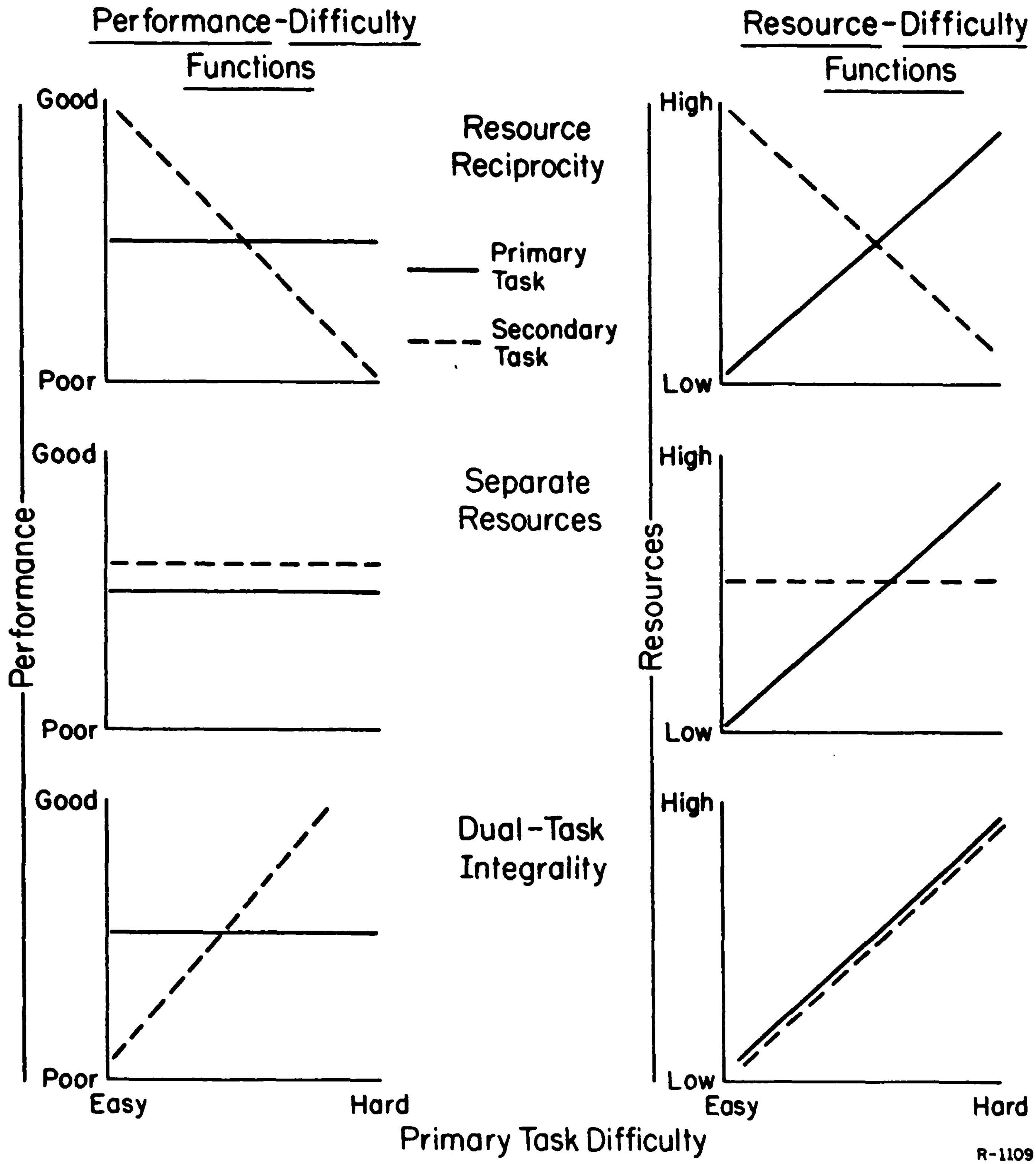


Spatial Discrimination



PURSUIT STEP TRACKING TASK




Performance-Difficulty Functions
Resource-Difficulty Functions
Good
Poor
High
Low
Resource Reciprocity
Primary Task
Secondary Task
Separate Resources
Dual-Task Integrality
Performance
Resources
Easy
Hard
Primary Task Difficulty
R-1109

during the performance of the pursuit step tracking task.

Figure 9. Grand average parietal ERPs elicited by the probes during the performance of the pursuit step tracking task in the first session. Subjects were instructed to ignore the probes in this session.

Figure 10. Grand average parietal ERPs elicited by the secondary task probes in the correlated and uncorrelated dual-task conditions.

Figure 11. A graphic summary of the P300 results. The amplitude of the P300s as a function of primary task difficulty is reported for each of the experimental manipulations. R represents the correlated conditions.

Figure 12. A model of dual-task integrality inferred from changes in the amplitude of the P300 as a function of primary task difficulty. The subscripted letters represent stimulus properties. The shade of the property lines represent the amount of processing. Both processing priority assignments (stimulus properties and tasks) are inferred from the amplitude of the P300 component. In the stimulus property case, some properties are processed to a greater extent then other properties (elicit a larger P300). The priority assignment for tasks is based on the relationship of P300 amplitude to the difficulty of the primary task. When properties are associated with the primary task objects, P300s increase in amplitude with increases in the difficulty of the primary task. On the other hand, when properties are associated with a separate secondary task, P300 decrease in amplitude with increases in primary task difficulty.

Figure Captions

Figure 1. The left panel presents the performance-difficulty functions. The right panel presents the corresponding resource-difficulty functions. Primary task difficulty is represented on the abscissa on both panels. The primary task is indicated by the solid line. The secondary task is represented by the dashed line.

Figure 2. A graphic representation of the pursuit step tracking task. The subjects task was to track the computer controlled target with the cursor along the horizontal axis. The difficulty of the tracking task was manipulated by changing the control dynamics from a first order to a second order system.

Figure 3. A graphic illustration of the spatial discrimination secondary tasks and the tracking task. The relationship of the task configurations to the experimental manipulations is represented along the abscissa (task display position) and the ordinate (relevant objects). In the same object condition, one of the tracking elements is also used for the secondary task discrimination.

Figure 4. A graphic illustration of the intensity discrimination secondary tasks and their relationship to experimental manipulations.

Figure 5. RMS error (a) and subjective difficulty ratings (b) for each level of system order during dual-task performance.

Figure 6. Grand average parietal ERPs elicited by changes in the spatial position of the tracking target in the dual-task blocks.

Figure 7. Varimax rotated component loadings for the first three components extracted in the Prinicipal Components analysis of the ERPs.

Figure 8. Grand average parietal ERPs elicited by the secondary task probes

Author Notes

This research was supported by the Air Force Office of Scientific Research under contract F49620-79-C-0233 with Dr. Alfred Fregly as technical monitor. We gratefully acknowledge the helpful comments of Michael Coles, Gabriele Gratton and Demetrios Karis.

Requests for reprints should be sent to Arthur F. Kramer, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel, Champaign, Illinois, 61820.

Engineering Psychology Laboratory Technical Report
EPL-83-5/ONR-83-5.

Wickens, C.D., Derrick, W.D., Micallizi, J. and Beringer, D. (1980). The structure of processing resources. In the _Proceedings of the 24th Annual Meeting of the Human Factors Society_. Santa Monica, California.

Wickens, C.D. and Kessel. C. (1979). The effects of participatory mode and task workload on the detection of dynamic system failures. _IEEE Transactions on Systems, Man and Cybernetics_, SMC-9.

Wickens, C., Kramer, A., Vanasse, L. and Donchin, E. (1983). The performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resource. _Science_, _221_, 1080-1082.

Wickens, C.D., Sandry, D.L. and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. _Human Factors_, _25_, 227-248.

Journal of Experimental Psychology, 18, 643-662.

Treisman, A. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. Perception and Psychophysics, 22, 1-11.

Trumbo, D., Noble, M. and Swink, J. (1967). Secondary task interference in the performance of tracking tasks. Journal of Experimental Psychology, 73, 232-240.

Vidulich, M. and Wickens, C.D. (1981). Time-sharing manual control and memory search: The joint effects of input and output modality competition, priorities and control order. EPL-81-4/ONR-81-4. University of Illinois at Urbana-Champaign, Engineering Research Laboratory.

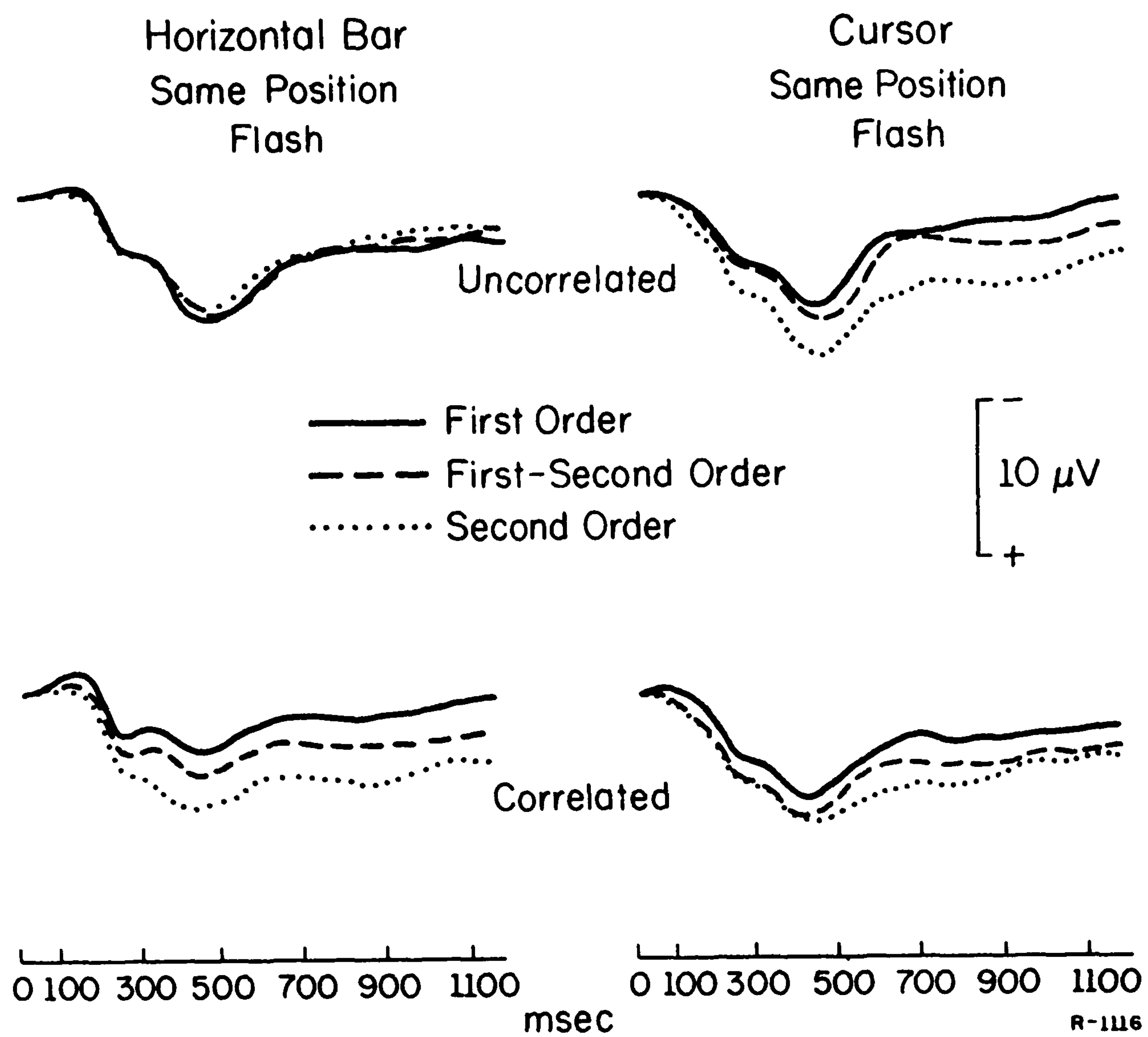
Weisstein, N. and Harris, C.S. (1974). Visual detection of line segments: An object superiority effect. Science, 186, 752-755.

Wickens, C.D. (1979). Human workload measurement. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press.

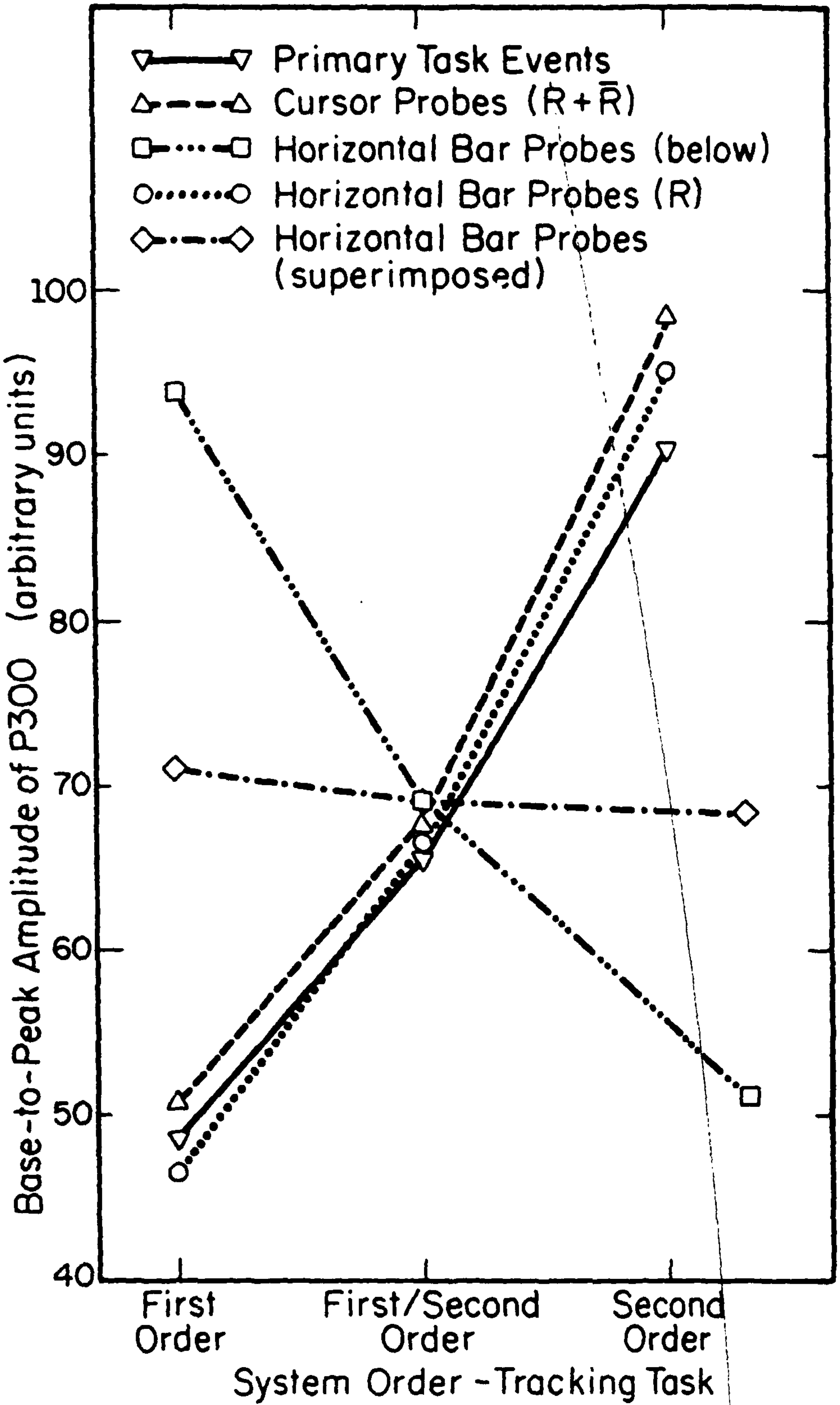
Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and Performance VIII (pp. 239-254). Hillside, New Jersey: Erlbaum.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

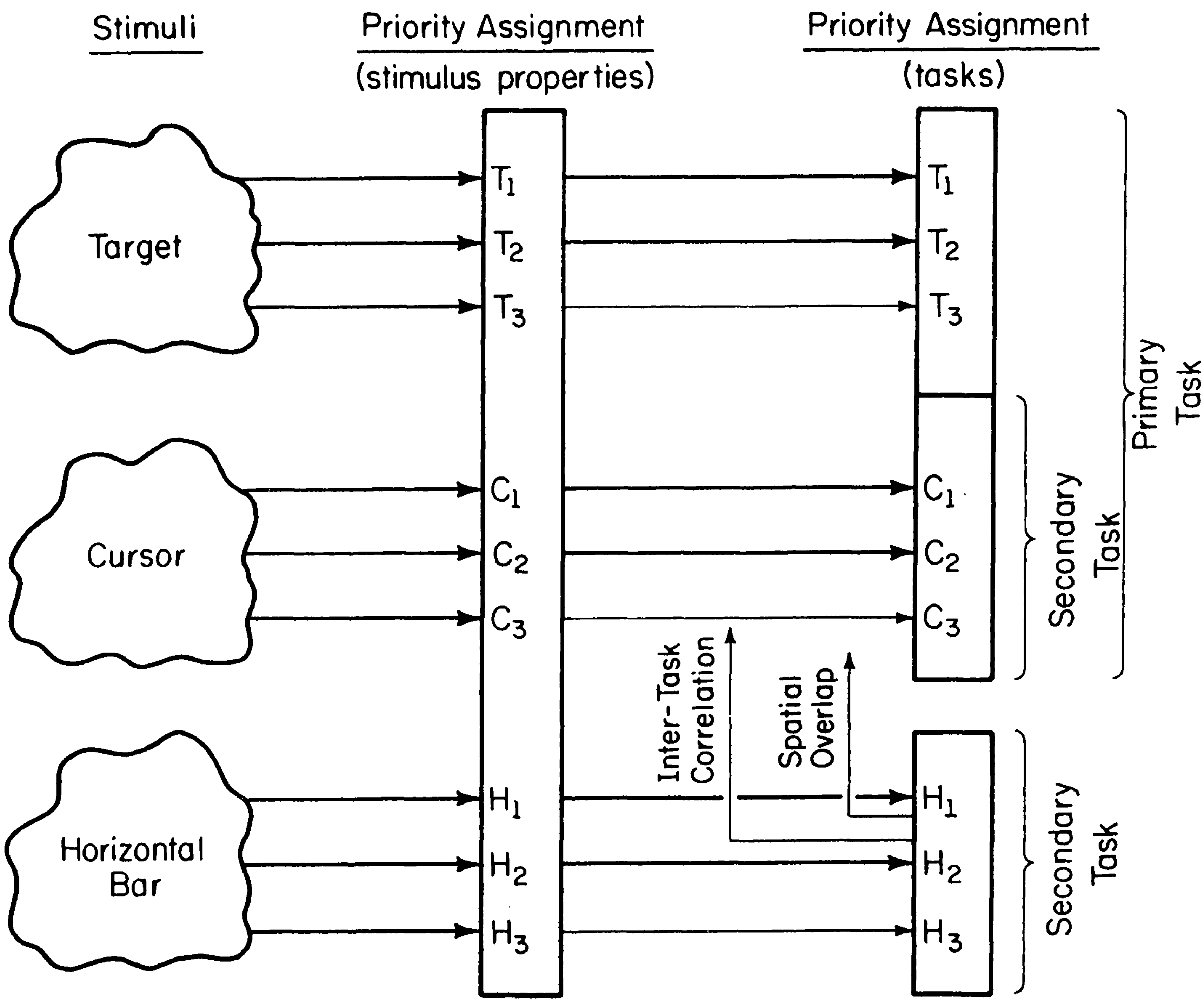
Wickens, C.D. and Boles, D.B. (1983). The limits of multiple resource theory: The role of task correlation/integration in optimal display formatting. University of Illinois.



Horizontal Bar
Same Position
Flash
Cursor
Same Position
Flash
Uncorrelated
First Order
First-Second Order
Second Order
10 μV
−
+
Correlated
0 100 300 500 700 900 1100
0 100 300 500 700 900 1100
msec
R-1116



Primary Task Events
Cursor Probes ($\dot{R}+\bar{R}$)
Horizontal Bar Probes (below)
Horizontal Bar Probes (R)
Horizontal Bar Probes (superimposed)
Base-to-Peak Amplitude of P300 (arbitrary units)
100
90
80
70
60
50
40
First Order
First/Second Order
Second Order
System Order - Tracking Task



Stimuli
Priority Assignment
(stimulus properties)
Priority Assignment
(tasks)
Target
Cursor
Horizontal Bar
T_1
T_2
T_3
C_1
C_2
C_3
H_1
H_2
H_3
T_1
T_2
T_3
C_1
C_2
C_3
H_1
H_2
H_3
Inter-Task Correlation
Spatial Overlap
Primary Task
Secondary Task
Secondary Task

APPENDIX B8

Flies in the Ointment:

The Use of P300 in Mental Chronometry

Michael G. H. Coles, Gabriele Gratton, & Emanuel Donchin

Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England, 1984.

Flies in the Ointment:

The Use of P300 in Mental Chronometry

Michael G. H. Coles, Gabriele Gratton, & Emanuel Donchin

Cognitive Psychophysiology Laboratory

University of Illinois

In this presentation, we examine the use of the latency of the P300 in mental chronometry. We first consider a number of objections that have been raised about our interpretation of the meaning of the latency of P300. We then report the results of a study illustrating the manner in which measures of P300 latency, coupled with those of reaction time (RT) and motor activity, can be a source of information regarding the temporal characteristics of mental processes. We will argue that the insight into *mental chronometry*, provided in this case by the P300, was not readily available from an analysis of more traditional measures.

The chronometric applications of P300 latency are based on the hypothesis articulated by Donchin and his co-workers (Kutas et al., 1977; McCarthy & Donchin, 1981) that the latency of the P300 is proportional to the duration of a subset of the information processing activities that follow an eliciting event. This duration may be shorter, or longer, than the duration of the subset of processes that determine the time at which an overt response is executed (the RT response). Evidence obtained in this laboratory and elsewhere (e.g., Duncan-Johnson & Kopell, 1982) suggests that those processes whose duration determines P300 latency do not include processes related to the selection and execution of the overt response. The class of manipulations that seem to have the strongest effect on the latency of the P300 have led to the suggestion that P300 latency is a measure of "stimulus evaluation" time.

A measure of the duration of stimulus evaluation processes, especially if it is not contaminated by response selection and execution, is clearly useful in the analysis of mental chronometry. Although other ERP components, such as N200, are excellent candidates for chronometric applications (e.g., Ritter, et al., 1983; Renault, 1983) the P300 has proven to be particularly useful. Its large amplitude, and the ease with which it is elicited, make it possible to study its latency on a trial-to-trial basis.

While our interpretation of P300 has proved to be attractive to those who would use it as a panacea for the problems of mental chronometry, it has been questioned on a number of grounds. Several of these "flies in the ointment" will be reviewed here.

1. "Noise manipulations do not increase P300 latency; rather, they give rise to new components."

This comment refers to the fact that in McCarthy and Donchin (1981), the noise condition was associated with two positive peaks, while only one peak was evident in the no-noise condition. We will argue that the results of a replication and extension of the McCarthy and Donchin study by Magliero et al. (1984) support the contention that the latency of P300 increases with noise. Specifically, Magliero et al. demonstrated that (a) the amplitude of the component elicited in the noise condition varies with task relevance/probability, and (b) that graded changes in noise are associated with graded changes in the latency of the component. In both cases, the P300 had a "classic" scalp distribution.

2. "P300 latency cannot be related to stimulus evaluation time since RT is sometimes shorter than the latency of P300."

This critique is based on the assumption that the information processing system consists of discrete stages arranged serially. Serial models of this type have been questioned recently by those who have proposed "cascade" and other parallel processing models. Once one accepts that many processes can act in parallel and that any of these can have an output at any time, the observed relation between P300 latency and RT is not surprising.

3. "Experimental manipulations that affect stimulus evaluation time have a larger effect on RT than they do on P300 latency."

This observation has been made by several investigators including McCarthy and Donchin (1981) and Ford et al. (1979). It is not inconsistent with our view of P300 latency, if we assume that manipulations of this kind have at least two effects. They influence the duration of those processes on which P300 depends (such as stimulus evluation time); but they also influence subsequent processes. Recent evidence from our lab suggests that response competition is one of these subsequent processes (Coles et al., in preparation).

4. "P300 latency shows small, but consistent, changes due to variables known to influence response selection".

Small variations in P300 latency as a function of "compatibility" manipulations have been reported by McCarthy and Donchin (1981), Ragot (1984), and Magliero et al. (1984). Stimuli delivering incompatible information may require additional processing compared to the same stimuli when they deliver compatible information. That is, strategic changes in the evaluation of the stimulus may occur as a function of compatibility

manipulations. Several experiments have shown that the P300 latency associated with a particular stimulus varies as a function of the context in which the stimulus is presented.

5. "On error trials, P300 latency may be affected by both stimulus and response evaluation."

This fly finds itself in the ointment because the evidence indicates that when the subject executes a fast guess in a choice RT experiment, P300 latency is very long. Note that only a single, delayed, P300 is observed on such error trials. We shall review evidence that suggests that in some circumstances the delay is due to an evaluation of the consequences of the subject's response. Indeed, the amplitude of the P300 elicited on such trials can be shown to predict the subject's performance on subsequent trials. In other circumstances, especially when the experimental conditions do not introduce a strong response bias, the delay on error trials appears to be related to the rate at which the information regarding the stimulus is accumulated.

In none of these objections do we find enough merit to suggest that the use of P300 as a measure of "stimulus evaluation" should be abandoned. The manner in which such studies can proceed will be illustrated by examining the P300 elicited in a letter recognition paradigm that has been studied extensively by Eriksen and his associates (1979). We will show that measures of P300 latency, when used in conjunction with measures of "behavior", provide support for a "continuous flow" model of information processing. The data also illuminate the nature of the evaluation process used by the subjects.

APPENDIX B9

The Use of ERPs to Monitor Non-Conscious Mentation

Emanuel Donchin

Paper presented at the 20th Annual Conference on Manual Control, Ames, California, 1984.

THE USE OF ERPS TO MONITOR NON-CONSCIOUS MENTATION

by Emanuel Donchin

Department of Psychology
University of Illinois

1. Introduction

1.1 The Washington Post Article

On June 3, 1984 the Washington Post carried an article by correspondent Michael Schrage entitled "Technology Could Let Bosses Read Minds." The article continued on the following page under the headline "Privacy Veil May Block Brain Watchers View." In the article Schrage reports that "Researchers in both academia and industry say it is now possible to envision a marketable product that could instantaneously assess whether employees are concentrating on their jobs by analyzing their brain waves as they work." Westinghouse's Research and Development center in Pittsburgh is described as "exploring the use of brain wave analysis - particularly a brain wave known as the P300 - as a means of determining an individual's level of attention and cognitive processing." The manager of the Human Sciences Laboratory in that Center predicts that "within the next 10 years Westinghouse could market a complete system capable of monitoring the mental processing effort of employees as they worked." The article goes on to review the opinions of others who are involved in the study of P300 lending to exchanges with one labor leader and one legal scholar, from Harvard, regarding the degree to which the use of P300 for reading the mind constitutes an "invasion of privacy."

The claims discussed in Schrage's article, and the worries they engender, have appeared frequently, in the past few years, in the public press and in scientific communications. The claims, and the concerns, are triggered by a solid body of evidence accumulated in several laboratories in the two decades since Sutton and his colleagues discovered the P300 (Sutton, Braren, Zubin, & John, 1965). The evidence suggests that the "endogenous" components of the Event Related Brain Potentials (ERP), and in particular the P300, can indeed be used as a tool in the study of cognitive function (Donchin, 1979). Indeed, much of this research has been supported by government agencies specifically in order to determine if it is possible to monitor, by means of the ERP, the operators of complex man-machine systems. The evidence does indicate that the ERP can provide data on aspects of the interaction between operator and task that may otherwise be opaque to monitoring (Donchin, Coles & Gratton, 1984; Kramer, Wickens and Donchin, 1983; Wickens, Kramer, Vanasse, & Donchin, 1983; Isreal, Chesney, Wickens, & Donchin, 1980).

Yet, it must be emphasized that these conclusions have yet to be tested in the crucible of practical applications. In the main, no research has yet been done to translate the laboratory findings into instruments that can be used by design engineers and by system managers. This is due, in part, to budgetary and to practical considerations. However the reluctance to invest in the development of ERP based monitoring may also be due to concerns regarding the appropriateness of using brain-waves to monitor mental activity. It is important therefore to emphasize that the "mind reading" implications of this work are often stated in a misleading and an inflated manner. We can indeed monitor mentation using the ERP. Furthermore, as I will endeavor to show in this paper, the ERPs provide a unique opportunity to monitor non-conscious mentation. Yet, it is not possible, and I believe it will never be possible, to use the ERP to "read minds" in the popular, friday night horror movie, sense of the phrase. My purpose in this lecture is to describe the class of inferences that can be based on ERP data and to emphasize the limits of these inferences. This, however, will not be an exhaustive review of the use of ERPs in Engineering Psychology. Rather, the application, its scope, and its limitations will be illustrated by means of one example. I will precede this example by a brief technical introduction to the methodology used in the study of ERPs.

1.2 Signal Averaging

Event Related Brain Potentials (or ERPs) are extracted from the EEG that can be recorded between a pair of electrodes placed on a person's scalp. The EEG is recorded as a continual fluctuation in voltage. It is the result of the integration of the potential fields generated by a multitude of neuronal ensembles that are active as the brain goes about its business. Within this "ongoing" signal it is possible to distinguish voltage fluctuations that are triggered in neural structures by the occurrence of specific events. This activity, evoked as it is by an external event, is known as the Evoked, or Event Related, Potential. It is but a faint whisper in the polyneural roar of the EEG. However, this whisper tends to follow the same time course whenever its eliciting event occurs. Therefore, when the EEG immediately following an event is examined over an ensemble of records the whispering ERP's are synchronized and their voice, as it were, becomes audible over the conflicting and asynchronous babble of the remaining EEG. Signal averaging is a technique for extracting such faint signals that follow a fixed time course relative to a trigger point. Detailed descriptions of the procedure are readily available (see Halliday, 1982).

The ERP extracted in this fashion takes the form of a series of fluctuations of the voltage between the recording electrodes. The epoch over which an ERP can be observed is on the order of several hundred milliseconds. The ERP is commonly considered to be a sequence of relatively independent components (Donchin, Ritter & McCallum, 1978). The amplitude of the components, their latency and their scalp distribution are the attributes of the ERP that are most commonly used in monitoring brain, and by implication cognitive, activity. Some of the components of the ERP, in particular those that appear within the first 100 msec following the stimulus are manifestation of the transformation, and the communication, of information in the sensory pathways. These "exogenous" components are generally followed by one or more components whose appearance, and patterns

of change, vary with the information processing demands placed on the subject. It is these, endogenous, activities that are used in monitoring cognitive activity.

1.3 How Are The ERPs Used in the Study of Cognition?

The monitoring tool to which Schrage's article refers is a record of a voltage change that can be obtained from the scalp of an awake human. These recordings can be obtained rather reliably and our knowledge has advanced to the point that we can predict with relative ease how attributes of these waves will change as a consequence of a variety of experimental manipulations. One readily obtained component is called P300, because it is positive going and its latency is hardly ever less than 300 msec. The P300 is often obtained in the so-called "oddball" paradigm in which a series of stimuli is presented to the subjects; the stimuli can be classified into two categories. If the events in one of the categories occur only rarely, then the rare events elicit an ERP that is characterized by a large, positive going, voltage change that peaks about 300 msec after the eliciting event. This late positivity is the P300 (see Pritchard, 1981; Donchin, 1981; Hillyard & Kutas, 1983 for reviews of the literature).

The P300, and other ERP components, provide an investigator with a set of dependent variables that can be used in the study of cognition. The manner in which these dependent variables relate to various independent variables is well established. However, as yet very little is known about the origin and functional significance of these signals. Evidence regarding the intracranial sources of the potentials is just beginning to emerge. It is likely that the ERPs represent the summation of potential fields associated with individual neurons who, fortuitously, are so oriented that their fields summate. But, as far as we can tell, the summated fields have no functional role in and of themeselves.

The nature of the ERP and the constraints on the interpretation of its physiological significance raises, inevitably, doubts regarding the validity and the utility of inferences made on the basis of these signals. Even though the Press has proven rather sanguine about the promise of ERP in monitoring cognition, the enthusiasm for its use has not proven infectious. Indeed, those who are most in need of techniques for monitoring the operators of complex systems have not been quick to adopt ERPs despite the very strong laboratory evidence for their utility. In part, this reluctance derives from a misunderstanding. It is commonly assumed that to be useful a "physiological" index must be directly involved in the processing activity being monitored. But this, I argue, is not necessarily a valid approach. In fact it is quite possible to conceive of a situation in which "epiphenomenal" indices may prove quite useful.

1.4 The Espionage Metaphor

The process by which the ERPs are utilized, its powers and its limitations, may be clarified by resorting to an analogy. I derive the analogy from electronic snooping. It seems that the design of computers is nurturing a new form of industrial espionage. These high-tech snoops record radiation emitted in the neighborhood of computing devices. It so happens that the structure of electronic data processing devices causes some of the

radiation emitted into the environment to be a manifestation of activity internal to the computer. Moreover, it is apparently possible to extract from this radiation, by appropriate computer analysis, useful data about the informational transactions that take place inside the computer. It is as if the information communicated within the computer's functional elements modulates recordable electrical activity in a manner that allows the perspicacious and enterprising spy to "read the mind" of the computer.

It is noteworthy that the activity recorded, and read, by such a spy is not necessarily a meaningful component from the point of view of the computer's information processing activities. The radiation may very well be due to the manner in which the computer was implemented. The availability of these extraneous signals depends on such factors as the choice of components and their packaging, the quality of the shielding. These are factors that are essentially irrelevant to the operation of the computer as an information processing device. Yet, however epiphenomenal, these activities that are "noise" to the computer are very much "signal" to the spy. Provided the technology for extracting the signals exists. Of course, such an indirect method will be used only if more direct methods to access the information of interest are not readily available.

I tend to view the ERPs in much the same way. For reasons having to do with the manner in which the brain is implemented, some of its activities are manifested on the scalp by a voltage change. It is likely that such activity is seen when many neurons are activated in synchrony and the topography with which these neurons are packed is conducive to the summation of their individual fields (Allison, in press). We assume that under the appropriate circumstances and with the appropriate analysis it may be possible to extract from these signals data that help in interpreting the activity of the brain. This is so because, as with the electronic spies, the actual informational transactions that take place within the brain modulate the ERPs, epiphenomenal as they may be, in ways that allow strong inferences about these informational transactions.

Note that, as with the extraneous radiation in the computer, we need not assume that the ERPs in themselves constitute a functional entity in the information processing executed by the brain. All we need to assume is that the intracranial entities that are manifested by the ERP play a role in information processing and that the modulation of the ERP, as the entities it manifests go about their business, is related in a systematic function to the activity of interest. With these assumptions we can observe variations in the ERP and draw inferences regarding the information processing activity. It is these inferences that allow the use of the ERP as a tool in the study of cognitive function.

To illustrate the manner in which the ERPs can be utilized, I will summarize a study by Gratton, Dupree, Coles and Donchin (in preparation) in which variations in the latency of one component of the ERP, the P300, has been used to reveal aspects of processing that accompany the responses of a subject who is performing an oddball task. The key assertion supported by this study is that ERP data can be useful in the examination of processes that are not readily available to introspection. By making the covert overt the ERPs can help in the study of non-conscious processes.

2. The Oddball Paradigm - Using Names

The study discussed here is one in a series of studies employing the Oddball paradigm in which the stimuli were names of individuals commonly used in the American culture. In all cases the series were constructed so that 20% (or, on occasion, 10%) of the names were names of males, (e.g., Jack, John, Eric...). All other names were names commonly associated with females, (e.g., Mary, Vanessa...). On some occasions, the subject was required to count the number of names that fell in one or another category, (a COUNT condition). On other occasions the subject indicated the occurrence of one of the categories by pressing one of two buttons, (a Reaction Time, or RT, condition).

The initial study in this series was reported by Kutas, McCarthy and Donchin (1977). Their subjects were presented with 3 different Oddball series. A "Variable Names" series was constructed from names of males and females as described in the previous paragraph. A "Fixed Names" series included just the names DAVID and NANCY. The third series was a sequence of words, 20% of which were synonyms of "PROD." The subject's task was to press one button in response to such synonyms and to press another button in response to all other words. The rare events in each series elicited a large P300. This was true regardless of the specific task assigned to the subject.

It turned out that the latency of the P300 varied across the 3 conditions. This was particularly noteworthy when the subjects were instructed to be accurate. The shortest latency was observed when the subject discriminated between the two names, David and Nancy. A longer latency is seen when the names vary from trial to trial. The longest latency was associated with the need to decide whether each of a rather disparate list of words is a synonym of PROD. These, and a considerable amount of additional data, lead us to suggest that the latency of the P300 depends on the time required for the evaluation of the stimulus. Subsequent work (McCarthy & Donchin, 1981), demonstrated that the latency of P300 is largely independent of the duration of processes that are involved in the selection and execution of the response. The interesting conclusion from these data has been that the latency of P300 is proportional to the time it takes to categorize the stimuli. If this is the case, the P300 latency may be used as a tool in mental chronometry to measure mental timing uncontaminated by "motor" processes (McCarthy & Donchin, 1983; Donchin, 1981). For studies in which P300 latency is indeed utilized in this fashion see Ford, Mohs, Pfefferbaum and Kopell (1980), Duncan-Johnson and Donchin, (1981), Goodin, Squires, and Starr (1983), Pfefferbaum, Ford, Johnson, Wenegrat, and Kopell (1983), as well as Coles, Gratton, Bashore, Eriksen and Donchin (in preparation).

2.1 The Correlation Between P300 Latency and RT

In a more detailed analysis of the data reported by Kutas et al. (1977), McCarthy and Donchin (1979) examined the relationship between the latency of P300 and the Reaction Time associated with each of the trials in an oddball study using names, sorted according to gender. The analysis capitalized on a filtering technique that allowed the measurement of the latency of P300 on individual trials (Woody, 1967). The principal finding

McCarthy, G. & Donchin, E. Event-related potentials: Manifestations of cognitive activity. In F. Hoffmeister and C. Muller (Eds.), Bayer-Symposium VII, Brain Function in Old Age. New York: Springer-Verlag, 1979, pp. 318-335.

McCarthy, G., & Donchin, E. A metric for thought: A comparison of P300 latency and reaction time. Science, 1981, 211, 77-80.

McCarthy, G., & Donchin, E. Chronometric analyses of human information processing. In A.W.K. Gaillard and W. Ritter (Eds.), Tutorials in event-related potential research: Endogenous components. Advances in Psychology, Vol. 10, G.E. Stelmach and P.A. Voon (Eds.). Amsterdam: North Holland Publishing Co., 1983, pp. 258-268.

Pfefferbaum, A., Ford, J.M., Johnson, R., Wenegrat, B., & Kopell, B.S. Manipulation of P300 latency: Speed vs. accuracy instructions. Electroencephalography and Clinical Neurophysiology, 1983, 55, 188-197.

Pritchard, W.S. The Psychophysiology of P300. Psychological Bulletin, 1981, 89, 506-540

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science, 1976, 193, 1142-1146.

Sutton, S., Braren, M., Zubin, J., & John, E.R. Evoked-potential correlates of stimulus uncertainty. Science, 1965, 150, 1187-1188.

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. The performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resources. Science, 1983, 221, 1080-1082.

Woody, C.D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical and Biological Engineering, 1967, 5, 539-553

Ford, J.M., Mohs, R.C., Pfefferbaum, A., & Kopell, B.S. On the utility of P300 and RT for studying cognitive processes. In Kornhuber, H.H. & Deecke, L. (Eds.), Motivation, motor and sensory processes of the brain: Electrical potentials, behavior and clinical use. Progress in brain research, Vol. 54, 1980, Elsevier/North Holland: Amsterdam.

Goodin, D.S., Squires, K.C., & Starr, A. Variations in early and late event-related components of the auditory evoked potential with task difficulty. Electroencephalography and Clinical Neurophysiology, 1983, 55, 680-686.

Gratton, G., Dupree, D., Coles, M.G. H., & Donchin, E. An electrophysiological manifestation of an error processing routine. In preparation.

Halliday, A.M. (Ed) Evoked Potentials in Clinical Testing. Churchill Livingstone: London, 1982, pp 575.

Hillyard, S., & Kutas, M. Electrophysiology and cognitive processing. Annual Review of Psychology, 1983, 34, 33-61.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 1980, 17, 259-273.

Johnson, R. E., Jr., & Donchin, E. On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography & Clinical Neurophysiology, 1978, 44, 424-437.

Johnson, R. Jr. & Donchin, E. Sequential expectancies and decision making in a changing environment: An electrophysiological approach. Psychophysiology, 1982, 19, 183-200.

Karis, D., Fabiani, M., & Donchin, E. P300 and memory: Individual differences in the von Restorff effect. Cognitive Psychology, 1984, 16, 177-216.

Klein, M., Coles, M.G.H., & Donchin, E. People with absolute pitch process tones without producing a P300. Science, 1984, 223, 1306-1309.

Kramer, A.F., Wickens, C.D. & Donchin, E. Analysis of the processing requirements of a complex perceptual-motor task. Human Factors, 1983, 25(6), 597-621.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197, 792-795.

McCarthy, G. Stimulus evaluation time and P300 latency. In E. Donchin (Ed.), Cognitive Psychophysiology Proceedings of Carmel I, Erlbaum, in press.

References

Allison, T. Recording and interpreting event-related potentials. In Donchin, E. (Ed.), Cognitive Psychophysiology Proceedings of Carmel I. Erlbaum, in press.

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W., & Donchin, E. A psychophysiological approach to the continuous flow model of cognitive processes, in preparation.

Donchin, E. Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked potentials and behavior. New York: Plenum Press, 1979, pp. 13-75.

Donchin, E. Surprise! . . . Surprise? Psychophysiology, 1981, 18, 493-513.

Donchin, E., & Bashore, T. Clinical versus psychophysiological paradigms in the study of event-related brain potentials. In S. Harnad (Ed.) The behavioral and brain sciences, in press.

Donchin, E., Coles, M.G.H., & Gratton, G. Cognitive psychophysiology and preparatory processes: A case study. In Kornblum, S.N. and Requin, J. (Eds.), Preparatory States and Processes, Hillsdale, NJ: Erlbaum Associates, 1984, 155-178.

Donchin, E., Kramer, A., & Wickens, C. Probing the cognitive infrastructure with event- related brain potentials. In Frazier, M.C., & Crowbee, R.B. (Eds.). Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics, AFFTC-JR-82-5, Edwards Air Force Base, 1982, 371-387.

Donchin, E., Kubovy, M., Kutas, M., Johnson, R., Jr., & Herning, R.I. Graded changes in evoked response (P300) amplitude as a function of cognitive activity. Perception and Psychophysics, 1973, 14, 319-324.

Donchin, E., McCarthy, G., Kutas, M., & Ritter, W. Event Related Brain Potentials in the study of consciousness. In Davidson, R. Schwartz, G. and Shapiro, D. (Eds) Consciousness and Self Regulation, Vol 3, Plenum Press, 1983, pp 81-121.

Donchin, E., Ritter, W., & McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), Brain event-related potentials in man. New York: Academic Press, 1978, pp. 349-441.

Duncan-Johnson, C., & Donchin, E. The relation of P300 latency to reaction time as function of expectancy. In H. H. Kornhuber and L. Deecke (Eds.), Motivation, motor and sensory processes of the brain: Electrical potentials, behavior and clinical use. Progress in Brain Research. Amsterdam: Elsevier-North Holland, 1981, pp. 717-722.

constraints of its nature and it better be applied within contexts that justify its usage. The available literature defines the nature of the information about an operator that can be extracted from the ERP. Whether this information is of utility in any given situation depends on the degree to which the information be utilized. If, for example, a man-machine system is not adaptive then it is entirely wasteful to provide it with information on the shifts in the operator's level of attention. The very same information may be extremely valuable, and well worth the cost of data-acquisition, if the system within which it is obtained is capable of adjusting to the operator's level of attention. In other words, the Psychophysiologist can point the availability of the information and define the methods by which it can be acquired. It is for the engineer and system designer to determine if this information can improve system performance at a reasonable cost.

One, of course, cannot be sanguine about the matter. If Polygraphy (lie-detection) can be used as a case in point, we must admit that when a technology that is capable of commercial exploitation becomes available the potent mix of the unscrupulous and the gullible may generate a vast industry. Polygraphy, like ERP research, utilizes a reliable phenomenon. It capitalizes on the fact that emotional changes are manifested by a class of recordable bodily changes. The interpretation of these changes in any given situation requires skill and a very careful analysis of the psychological structure of the situation. It may, in very carefully designed tests, in the hands of well-trained, experienced, Psychophysiologists yield valuable information about the veracity of a witness. To move from this to the application of the polygraph in personnel offices to screen job applicants is bizarre indeed. I dearly hope that we shall not see in the near future the appearance of ERPgraphers, wielding Signal Averagers, assessing workers' productivity to the joy of gullible corporate managers.

The need to guard against the avaricious and the naive should not obscure the vast possibilities opened by Cognitive Psychophysiology for a better understanding of human performance, and for monitoring operators in useful ways. The P300, and the other ERP components, clearly provide useful data. Our knowledge of these signals is still in its earliest stages. I am confident, however, that the range of useful information that can be extracted from the ERP will be extended in the coming decades. There is already sufficient data to justify the incorporation of ERP measures in the design phase of complex systems. The closed-loop application that comes to mind when we consider monitoring an operator may be a thing of the remote future. However, the P300 can be of considerable use to designers who need to evaluate several competing systems in terms of the effectiveness with which operators can use the systems. The development effort, to my mind, should be devoted largely to the utilization of this valuable window on the mind in the design, rather than during the actual use, of Person-Machine systems.

an error trial, the more likely is the subject to be correct in the response the next time a male name is presented regardless of the number of female names that have appeared in the interim. It would appear that subjects indeed modulate their response bias when an error is discovered. More important is the observation that the degree to which this shift in strategy takes place is indexed by the P300. These data strongly support the proposition that the amplitude of the P300 reflects the intensity with which a context-updating process has operated.

That there is indeed a shift in the bias is supported by the analysis of the Reaction Times associated with the presentation of Female names that occurred immediately after an erroneous response was made to a male name. If the subject is indeed shifting response bias in the direction of Male names, we expect the responses to female names to be slowed down in the trials immediately following missed Male names. This increase in RT should be proportional to the amplitude of the P300 elicited on the error trial. This, is precisely what we found. The larger the P300 elicited on a given trial the slower is the response to the immediately following female names.

It is interesting that the latency of the P300, delayed as it may be, does not predict the response on subsequent trials. But, than, this should come as no surprise. The latency is index of the duration of the processes preceding the invocation of the P300. Thus, it is not directly related to the process which in fact updates the context. The latency should therefore should therefore have no effect on the subject's model of the environment. And indeed, we could detect no relationship between P300 latency and subsequent performance.

5. Conclusions

5.1 The Implications for Monitoring

The nature of the information on mentation that can be gleaned from ERPs is illustrated by the data I have just described. The study is quite typical in the evidence it yielded and in the complexity of the procedures required to interpret the evidence. How likely is it that devices for measuring the P300 will appear, let alone proliferate, in the work place in the coming decades? It seems clear that the ERPs do provide information that is not otherwise available. However, it should be equally clear that the language with which the ERPs speak is arcane. The significance of the presence, or absence, of a P300 and the interpretation of modulations of its amplitude and latency can be assessed only within the framework of a careful analysis of the circumstances. The amplitude of P300 can increase, or decrease, for a large number of different reasons. In a carefully structured situation the interpretation, to the trained and skilled investigator, is not too difficult. But, it is unlikely that it would be possible to attach a machine that would yield a simple, universal, situation-independent, number that can be used by a manager, a designer, or even the operator to make intelligent on-line decisions.

Of course, it is not my intention to suggest here that the efforts to develop the ERP as a tool for the Engineering Psychologist were wasted. I do believe that the ERP is a unique and valuable tool. However, it must be realized that, as is true for any tool, it is best used within the

model in that specific predictions can be derived from that model regarding the consequences of the P300. For example, Klein, Coles and Donchin (1984) have shown that people with perfect pitch process phonic probes without emitting a P300. That this would be the case was predicted on the basis of the context updating hypothesis. Karis, Fabiani and Donchin (1984) have shown than the amplitude of the P300 elicited by a stimulus in a study of the von Restorff effect predicts whether or not the stimulus will be recalled.

4.1 The Delayed P300 on Error Trials--An Interpretation

If the process manifested by the P300 performs a function that is necessary for the maintenance of the model of the environment in Working Memory than it may be suggested that it is not invoked until the data needed for determining the needed changes is available. We propose that the delay in the P300 on error trials is inserted as the error is recognized by the system because there is a need for further processing before the book can be closed on the trial. Note, than in our view the P300 process is invoked in order to serve the needs of action on future trials. Thus, the elicitation of P300 on when the rare stimulus appears may be associated with the resetting of the system to accommodate responses to the rare events. After all, the subject is clearly biased to emit the frequent response at the slightest provocation. One assumes that these responses are emitted as soon as the appearance of a stimulus is detected. As processing of the stimulus continues, after the response has been made, the name is properly encoded. The conflict between the category of the name and the response forces on the system additional processing. The additional time required for this processing is the delay we observe in the P300.

We are fairly confident that the delay in P300 on error trials is indeed associated with the recognition of the error. Though we emphasize that we are not implying that this is a conscious, intentional, delay. Other plausible alternatives have been considered and have been ruled out, (Gratton, et al., in preparation). The proposal is plausible. However, the plausibility does not provide adequate support for the theory. The critical test, again, is the ability to derive from our interpretations of the delay specific predictions. In this case, the proposal that the process manifested by P300 serves the responses made by the subject on future trials suggests that there ought to be a relationship between the amplitude of the P300 elicited on error trials and performance on succeeding trials. We conducted two such tests to evaluate the validity of this view.

4.2 The Amplitude of P300 on Error Trials And Its Consequences

If subjects err because they are biased to respond to the frequent event than one consequence of the recognition of an error would be an attempt to shift the bias away from the activation of the frequently pressed button. The shift would be in the direction of the response to the rare event. Such a shift should be accompanied by an increased probability that response will be given on the "male" button to male name. If the P300 is an index of the degree to which readjustments of the system's model of the environment than, the larger the P300 the large we would expect the shift to be. We examined therefore the subject's responses on all trials in which a Male name was presented. It turns out that the larger the P300 elicited on

held constant than P300 amplitude is determined by the extent to which the task with which the P300 is associated is at the focus of the suoject's attention. This indeed is the basis for the use of P300 as a measure of Workload (Isreal, et al., 1980; Kramer, et al., 1983; Donchin, Kramer & Wickens, 1982). It is also clear that while the rarity of the eliciting event can play an important role in the elicitation of the P300, rarity is neither a sufficient, nor a necessary, condition. Studies of P300 elicited when subjects are assigned dual tasks indicate that P300 is a manifestation of processes associated with perceptual, categorization, activities. In addition evidence has been presented that the amplitude of P300 is inversely proportional to the degree to which an earlier representation of the stimulus has decayed (Squires, Wickens, Squires & Donchin, 1976).

With these ensemble of antecedents on hand one can proceed to the next two stages of the theory building process. These data, if sufficiently complete can lead to a model of the P300 couched in the terms we required above. That is, a statement need be made that assigns a function to the P300. The statement represents an integration and an interpretation of all that we know about the P300's antecedent conditions. To be useful it is not sufficient for this model to be merely a plausible summary of the available data. Rather, it should serve as the basis for the third, the theory testing, phase. In that last phase predictions that are derived from the hypothesis we entertain regarding the component's function need be tested. Such predictions take the form of statements about the consequences of the P300.

As I argued elsewhere (Donchin, 1981), if the P300 is a manifestation of a processing entity, a subroutine if you will, than it must have outputs that feed into subsequent, or parallel, stages of the information processor. If the amplitude of the component is proportional to the intensity of its activation, than its activity will affect subsequent processing stages in a manner that is related to the amplitude of P300. In other words, it must have consequences. If we believe we know its function, we ought to be able to predict these consequences. It is in the generation and the testing of such hypotheses that theories regarding the P300 are tested.

4. A Hypothesis Regarding the P300

The specific hypothesis that currently serves as a guide for the work my colleagues and I are conducting at the Cognitive Psychophysiology Laboratory at the University of Illinois views the process manifested by the P300 as an instrument in the service of the operation of Working Memory. By this term we refer to the ensemble of representations that are, at any time, in a state of higher availability. The membership in this ensemble is continually changing as the needs of the moment change. For any given task, some new representations may be needed, while others (remaining from previous tasks) must be discarded. The process is dynamic and requires, one should assume, a considerable amount of housekeeping. There must be an ongoing process of context evaluation and context updating. I have argued that the P300 is a manifestation of a processing entity that is utilized while such context updating, or memory management, takes place.

Whether this model will ultimately prove to be a good approximation to the truth remains to be seen. However, it does satisfy the criteria for a

degree to which the functional significance of the component is known. In the specific case we are discussing here we need to have a theory regarding of the functional significance of the P300 so that a framework is available for assessing the implication of its increased latency.

How does one go about elucidating the functional significance of an ERP component? In my view a three fold process is required (see Donchin, 1981; Donchin & Bashore, in press; Donchin, et al., 1984). The entire process is guided by a view that sees an ERP component as the manifestation of an intracranial processor which implements some information processing operator. This statement raises some complex philosophical issues (see Donchin & Bashore, in press). However, in its simplest form the relation between the ERP and mentation is viewed in much the same form as are the radio emissions discussed in Section 1.4. The principal implication of this view is that theories regarding the functional significance of the ERP are best developed within some comprehensive model of information processing. The hypothesis regarding the component's function will be stated by identifying a processing element within the general model. Such an element is defined in terms of the transformations it performs on its input. A theory of the P300 than asserts that the component's appearance indicates that this particular operation has been invoked. The component's latency is a measure of the duration of processes whose occurrence must precede the invocation of the processor. The amplitude of the component is taken as a measure of the intensity with which the critical operation has been performed. Many assumptions are implicit in this description of theory building in Cognitive Psychophysiology. Some are more tenuous than others. Thus, inferences about the latency of a component are fairly straightforward. On the other hand, the interpretation of the amplitude as a measure of the utilization of the component (Donchin, Kubovy, Kutas, Johnson, & Herning, 1973) is based largely on faith, on the plausibility of the assumption and on the fact that this is as good a working hypothesis as we can muster.

3.3.3 The Need for Theory Testing

A theory of the P300 must begin with an enumeration of what I have called the antecedent conditions of the component (Donchin, 1981). In effect, the bulk of the research on P300, including the study described in detail in this lecture, has been concerned with the enumeration of these antecedent conditions. This search yields an ensemble of statements that describe the conditions under which the P300 is elicited. There is also a need to determine the functional relationship between variations in many aspects of the eliciting situation and attributes of the P300. Much effort has been invested in determining the factors that control the amplitude of the P300, its latency and the variation in its scalp distribution. Such data have accumulated in the last two decades to an extent that permits a rather precise enumeration of the antecedents of the P300.

3.3.4 The Antecedents of the P300

The list is familiar (Hillyard & Kutas, 1983; Pritchard, 1981). The P300 is elicited by rare, task relevant, events. If task relevance is held constant than the amplitude of P300 is inversely proportional to the subjective probability of the eliciting event. If subjective probability is

and incorrect trials. But, establishing the existence of such a difference is not a particularly satisfying enterprise. In the first place it is not all that surprising that such a difference is observed. Moreover, the existence of such differences has been established quite persuasively by means of the classical methods of Cognitive Psychology. What do we gain, how do we augment the available knowledge, by adding the ERP to our armamentarium?

3.3.1 The ERP and Non-Conscious Mentation

It would seem that one of the principal values of the ERPs is that they allow observation of processes that do not have obvious representations in awareness. That such processes exist goes almost without saying. We are not aware, and most probably can not be aware, of most of the internal information processing activities that yield as a consequence the contents of awareness. Consider Speech. By and large we are aware of the content of our discourse. We know what we say, we may know why we want to say what we say and we know the purpose underlying our words. These all are the contents of consciousness. Yet, we are at the same time entirely unaware of the nature of the process used to select our vocabulary, or sort out these words into proper grammatical sentences. Even when we consciously search for a word, we are blissfully unaware of the manner in which our mental gears grind as the word is searched for. When candidate words are dredged, we know immediately - we are fully "aware" of - the degree to which that word is, or is not, a suitable choice. But, if we know it is not the correct word, how come we cannot find the proper word? These processes, and much more that is of interest to the cognitive scientist, takes place well outside consciousness.

It is in fact these non-conscious activities that <u>are</u> the principal focus of interest to Psychologist. True, as persons we <u>are</u> principally interested in that of which we <u>are</u> aware. But, as Psychologists we are interested in the processess <u>underlying</u> the observed behavior. We would like to understand how memory is organized and how information in memory is searched and is retrieved. We would like to know how sensory information is integrated into the percepts of objects and how the speech stream is scanned into words whose meaning is extracted even as all their related associations are activated. These are the psychological operations whose elucidation is the goal of Cognitive Psychology. As these are largely non-conscious the Science is based on inferences from observations on the pattern of overt behavior. Alternately we depend on self-reports, a rich but occasionally flawed record. It seems that, at least to a limited extent, ERP components allow us to monitor directly the intensity and the latency of some of these processes (Johnson & Donchin, 1978; Johnson & Donchin, 1982; Donchin, McCarthy, Kutas, & Ritter, 1983).

3.3.2 The Research Design

But, even if one grants that the ERP is a manifestation of brain activity which implements an interesting mental operation, and hence by implication the ERP can be considered a manifestation of such mental operations, how does one determine the nature of the specific operations associated with a specific component. Clearly the degree to which the P300 or any other component could be used for monitoring operators depends on the

subject and that the performance on subsequent trials is affected by such processing. This error processing need not call on the subject's awareness. The error may be processed, and its consequences integrated into the response stream, whether or not the subject is conscious of the error. Indeed, the existence of error-related processing has heretofore been inferred from variations in the performance on trials that follow the error. An examination of the ERPs acquired by Gratton, et al. reveals that some intracranial processing entity is affected by the occurrence of an error.

Support for this claim is provided by examination of the ERPs elicited by names of Males and of Females. The EEG activity was sorted so that the ERPs associated with correctly identified and mis-identified Male names are plotted separately, as are the responses to Female names. The data were also sorted according to the speed of the response. The bottom panel plots the data from the fastest responses, each successive panel represents slower responses. The data are clear. The ERPs elicited by the missed Male names and by the Female names are quite different in pattern. The Male names elicit a substantial P300, the Female names barely do. Thus, the homogeneity of the motor responses obscures a difference between the activity of whatever intracranial system is manifested by the P300. As the response topography of the Male and Female responses appears to be quite similar, it is difficult to attribute the delay in the latency of names to the P300 to the speed with which the subjects respond on the error trials. The speed of the response on a Female trial is equal to the speed of the response on the incorrect Male trial.

There is also a patent difference between the ERPs elicited by Male trials that were correctly identified and those that were missed. The peak positivity on the error trials is delayed by almost 100 msec. This finding corroborates the reports by Kutas, et al. (1977). A detailed analysis of the distribution of the component supports the identification of the delayed component as the P300 (see Gratton, et al., in preparation). Thus, we confirm the paradoxical relationship between the RT and the latency of P300. The relatively short RT's associated with the incorrect trials are accompanied by a P300 with a long latency. Conversely, when the RT is relatively long, as it is on the correct trials, the P300 latency is short. It is important to note that this pattern of results holds for all the conditions used in this study. Error trials were associated with the longer latency P300s when the probability of names in the two categories was equal. Similarly, the result held when subjects tried for accuracy. Moreover, the pattern was maintained even when the data were sorted according to the speed of the response. That is, when trials are classified into bins according to the RT on each trial, then within <u>each</u> bin the error trials are associated with longer latency P300.

3.3 Interpretation

It seems, therefore, that it would be prudent to accept the empirical assertion that the P300 tends to have a substantially longer latency on trials on which the subject pressed the wrong button. How can we interpret such an observation? What, if anything, does it tell us about the mental activities that take place as the subject is performing the assigned task? The empirical statement, by itself, can support the conclusion that there is a difference of some sort between processing activities accompanying correct

to linked mastoids. EOG was recorded for purposes of subtracting out ocular artifact from EEG, with a Beckman electrode placed above and to the right of the right eye. EMG was recorded by two Beckman electrodes placed one half an inch apart, one third of the distance on the diagonal between the elbow and the outer wrist when palm up. Analog to digital conversion occurred for 1200 msec which consisted of 100 msec of baseline before each stimulus name and 2200 msec from the movement of presentation.

3.2 Results

A detailed presentation of the rather large amount of data, and the numerous analyses of these data will be given in Gratton, et al. (in preparation). Here, I shall summarize some of the results focusing on the data obtained when the Male names where rare an the subject was urged to be fast, (the "*speed*" *condition*). I will *not* present here the statistical analyses that support my various assertions. Again, these are presented with some detail by Gratton, et al. (in preparation). The reader can rest assured that all statements made here are backed by adequate analyses.

3.2.1 Reaction Time Data

3.2.1.1 Histograms for Individual Subjects

The pattern of Reaction Times was consistent. Subjects respond with virtually no errors to Female names. They do so rather fast. That is, the RTs associated with female names tend to be short and the number of errors, that is presses on the Male button in response to a Female name, is miniscule. The pattern for Male names is quite different. Correct responses to Male names are rare and, when given, they are given slowly. On the other hand, it is clear that on most trials on which a Male name is the stimulus the subject presses the "Female" button. Moreover, the RT on these trials tends to be quite short. The RT in this case is in fact quite similar to the RT associated with the correct Female name.

The data indicate that the subjects' responses differed according to the button they pressed, or the hand they were using. The Male button was pressed solely in response to the appearance of Male names. The speed with which these responses were made was always slower than was the speed of response on the Female button. It is plausible to assume that the subjects were primed to *respond with the hand* that was called upon to respond most frequently. This "response bias" caused the subject to respond on many a trial to the Male name with the response on the Female button. It is striking that the distribution of the RTs for these fast responses is rather independent of the eliciting stimulus. Pressing, correctly, for a Female name and committing a "fast guess," by pressing the same button in response to a Male name are indistinguishable as responses, at least as far as the shape of the distribution is concerned.

3.2.2 ERP Data

While the correct overt response is indistinguishable from the overt erroneous response the processing associated with the two classes of responses is likely to be quite different. There is considerable evidence that fast guesses, and other errors, are monitored and processed by the

In the present case, the claim that is in need of evaluation is that the P300 reveals, through modulations of its latency, the activation of an internal, mental, process that is invoked as a consequence of the recognition that an error has occurred. If we can be sure that the peak with the longer latency is indeed a delayed P300 rather than a new component, and if we can be sure that the delay is indeed due to the occurrence of the error rather than to such factors as the speed of the response associated with the movement, than the P300 is indeed revealing in a unique fashion aspects of the information processing system. To resolve some of the doubts that remained regarding the ERPs elicited on error trials we replicated, and extended, the study reported by McCarthy and Donchin (1979).

3. A Study of P300 Latency on Error Trials

Thus, we have again presented subjects with a series of names. In one series the names appeared with unequal probability, names of Females appearing frequently, P(female)=.80. In another experimental condition the two categories appeared with equal probability. These two probability conditions were crossed with two performance regimes. In one the subject was instructed to respond as fast as possible. In the other regime the subject was told to be as accurate as he could. From each of the 7 subjects we obtained 800 trials in each of the conditions.

3.1 Design

Procedure. The subject was positioned in front of a PLATO terminal with the fingers of each hand resting around a 2" diameter bar of a dynomometer. The choice-reaction-time task required a sharp squeeze and release of the bar from one hand in response to male names appearing on the screen and a squeeze and release from the other hand in response to female names. Names were presented one at a time in the center of the screen for 200 msec with a 2000 msec interstimulus interval. A list including 10 male names and 10 female names was used to generate the series. The four to seven character names were chosen for their familiarity and for the certainty of their gender.

Subjects were shown the names in blocks of 100 trials. Blocks were made up of either 80 females and 20 males or 50 of each. Also, subjects were instructed to respond as quickly as possible or as quickly as possible without making errors. The two conditions, (1) the relative probability of male and female names and (2) the instruction set (speed or accuracy), were factorially combined, resulting in eight experimental cells. Eight hundred trials were run in each cell, with half the trials run during one session and the remaining half run during a second session. During each session, four blocks of 100 trials were run for one experimental cell at a time. The order of conditions was counterbalanced across subjects in a latin square design, and the order of conditions run during the first session was reversed for the second session. Also, the relationship between the class of stimuli (male or female names) and the responding hand (left or right) was counterbalanced across subjects.

In addition to response time, EEG was recorded by Ag-Ag Cl electrodes at Fz, Cz, Pz, C1, and C2 placed according to the 10-20 system and referred

commission of the error. Several alternate explanations can be invoked. Two of these difficulties are summarized here.

2.3.1 New Component?

One of the major difficulties presented by ERP data is associated with the definition and the proper identification of components of the ERP. For example, each of the positive going peaks observed by Kutas et al. (1977) in the ERPs elicited by the three series has been labeled "P300" even though the peaks differ in latency by as much of 100 msec. What leads us to believe that these three peaks are indeed instances of a component whose latency is shifted by the duration of the processing precedes its invocation? How do we know that the peaks with the longer latencies are not entirely new components that are elicited by the presentation of a word, or by the search of a synonym. The issue is generally resolved on the basis of the similarity of wave shapes, on the scalp distribution of the potentials and on the manner in which they respond to experimental manipulations (Donchin, et al., 1978). There remains the possibility that delayed peaks that are recorded in association with error trials are different components rather than a delayed P300.

2.3.2 Response Related Factors

Another interpretation of these data is based on the fact that on all these error trials the subject responded rather fast to the stimulus. In other words, these are clearly trials on which a variety of factors are injected into the stream of processing. How do we know that it is the recognition of the error, rather than the fact that a very fast response was emitted on the trial that accounts for the delay? A different, but related possibility is that it is not that P300 is delayed on error trials, but rather that errors may be more likely on trials on which P300 latency is long.

2.3.3 The Need For an Additional Study

The controversy surrounding the interpretation of the ERPs recorded on error trials touches on some of the key issues in the interpretation of the ERP. The manner in which such controversies arise, and the action that is needed to resolve the issue, must be understood if these data are to be used in the, so-called, "real" world. Any monitoring system that utilizes ERPs in the manner described by the Washington Post article will, in one way or another, acquire data much like those described above. Essentially the data analysis, however sophisticated, boils down to a comparison of the amplitudes of waveform features obtained at different sites on the same occasion or features that were obtained from the same site on different occasions. Whenever such a comparison is made it is critical to assure that one compares features of the same object. If it is possible to mistake one component for another, then shifts in latency or in amplitude that are assumed to reflect shifts in the allocation of attention may in fact reflect an altogether different process. Such a confusion will frustrate any attempt to utilize the ERPs, regardless if the use is made in a laboratory or an industrial environment.

has been that the correlation between P300 latency and RT depends on the strategy adopted by the subjects. When the subjects were instructed to be accurate the correlation between P300 latency and RT was significantly different than zero. On the other hand, when instructed to be fast, the subjects' RTs and P300 latencies were quite uncorrelated. These data supported the suggestion (Donchin, 1979, 1981) that the P300, and the motor response, may each be the culmination of a series of processing activities and that these streams of processing can, in principle, be quite independent of each other.

The P300 latency is assumed to reflect the duration of stimulus evaluation processes. From the evidence on hand it would appear that the processes leading to the invocation of a P300 continue for as long as is required for a full evaluation of the stimulus. The latency of P300 is, therefore, at least as long as the duration of these evaluation processes. The overt responses, on the other hand, may well be released "prematurely" on the basis of limited information. The correlation between Reaction Time and the latency of the P300 will therefore depend on the degree to which the overt responses that define the RT are made contingent on the full evaluation of the stimulus. The more inclined the subject is to respond prematurely, the poorer the correlation between the latency of the P300 and the RT.

2.2 The P300 On Error Trials

One striking aspect of the data acquired by McCarthy and Donchin (1979) was observed when the trials on which subjects made errors. These were trials on which the subject responded to a rare event as if it was frequent. That is, even though a Male name appeared on the screen, the subject pressed the button associated with Female names. There were but a few such trials in the study reported by McCarthy and Donchin (1979). However, in virtually all these trials the pattern was the same - the Reaction Times were relatively short and the P300 latency was relatively long. It was as if on these trials the subjects first acted and than thought! As the number of error trials was small, we replicated the experiment presenting the subjects with many more trials and pressing even harder for fast responses. A partial report on these data can be seen in McCarthy (in press). In 10 out of the 11 subjects the pattern obtained was identical. Errors of commission, "fast Guesses," were associated with very short RTs and relatively long P300 latencies. McCarthy and Donchin (1979) suggested that whenever an error was detected on any given trial, the invocation of the P300 was delayed. The delay was required, presumably, to allow further processing of the trial's data. This interpretation exemplifies the manner in which observations of the P300 lead to inferences regarding an internal process even though these processes may not be readily observable by conventional means.

2.3 Puzzles for Present Experiment

Even though the increase in the latency of the P300 was quite evident in the data obtained by McCarthy and Donchin (1979) it was not sufficient to support the conclusion that this delay is due to extended processing consequent on an internal, not necessarily conscious, recognition of the

APPENDIX B10

Effects of Mnemonic Strategy Manipulation in a von Restorff Paradigm

Monica Fabiani, Demetrios Karis, and Emanuel Donchin

Paper presented at the Third International Conference on Cognitive Neuroscience, Bristol, England, 1984.

ICON III Conference - Bristol (England)

Abstract

Effects of mnemonic strategy manipulation in a von Restorff paradigm

Monica Fabiani, Demetrios Karis and Emanuel Donchin

Cognitive Psychophysiology Laboratory

University of Illinois at Urbana-Champaign

In a previous study (Karis, Fabiani & Donchin, 1984), we recorded ERPs to items in a series in which a deviant item ("isolate") was embedded. It is commonly found that isolates are better recalled than comparable non-deviant items (von Restorff effect). We found that subjects who displayed the largest von Restorff effect were also the worst in recalling list items. Furthermore, these subjects reported that they used rote strategies to memorize the words. For these subjects, isolates that were recalled elicited on initial presentation a larger P300 than was elicited by isolates that were not recalled. Subjects displaying the lowest von Restorff effect had the best recall performance and reported using elaborative strategies to aid their recall. For these subjects, P300 amplitude did not predict subsequent recall.

We ran a second experiment to test the reliability of this finding and to determine if it is indeed the case that the relationship between recall and P300 amplitude depends on the subject's recall strategies. Six female subjects participated in a von Restorff experiment and were given explicit strategy instructions. During two experimental sessions (in which ERPs were recorded) they were instructed to use either rote or elaborative strategies to memorize the words. The pattern of behavior we found for the same subjects operating under different instructions was similar to the behavior

we observed in different subjects in the previous experiment. When instructed to use rote strategies, subjects displayed a significantly higher von Restorff effect and a lower performance than when instructed to use elaborative strategies.

Preliminary analysis of ERP data for the two subjects showing the largest behavioral change from one strategy to the other supports our hypothesis that amplitude of P300 is related to recall only when the subjects are using rote strategies but not when they are using elaborative strategies.

Component identification with Vector Analysis

Gabriele Gratton, Michael G. H. Coles, and Emanuel Donchin

Cognitive Psychophysiology Laboratory

University of Illinois at Urbana-Champaign

Abstract of a paper to be presented at the III International Conference on Cognitive Neuroscience - session on Methodological Issues in Cognitive Electrophysiology. Bristol, England, September 17-21, 1984.

This presentation will be devoted to the illustration of a procedure, Vector Analysis, for the quantification of scalp distribution information. In the first part, the basic assumptions and formulations of Vector Analysis will be reviewed. Then two examples of the application of the procedure will be presented.

In contrast with most procedures which analyze scalp distribution, Vector Analysis considers the values observed at several electrode sites as the sum of several components, each characterized by a specific pattern of scalp distribution, and of background noise. The emphasis on the concept of "component" is one of the major aspects of Vector Analysis. It allows the investigator to address the problem of isolating and quantifying the contribution of particular components. A second major aspect is the adoption of a multivariate approach: different electrode locations are considered as different variates, which reveal different aspects of the same phenomena, and which share both signal and noise contributions.

The interest in scalp distribution is based on the assumption that the electrical activity of certain brain structures involved in psychological processes is manifested at the scalp by particular components. The scalp distribution of each component reflects anatomical and physiological properties of the structures involved in the generation of the component, as

well as conductive characteristics of the interposed media. Several, largely unsolved, problems are associated with the task of localizing the generating structures of ERP components by means of maps of scalp electrical activity. However, most investigators agree that ERP components are characterized by specific scalp distributions. This is probably due to the invariance of the underlying source(s). This observation has led several investigators (see Donchin, Ritter, and McCallum, 1978) to consider scalp distribution as one of the defining characteristics of an ERP component. Hence, the use of information about scalp distribution may be helpful in the identification and analysis of ERP components.

The basic assumption of Vector Analysis is that the potentials recorded at several scalp electrodes are given by the sum of several ERP components and of noise. Each ERP component is characterized by a specific scalp distribution. This specific scalp distribution may be expressed by a series of weights, one for each electrode. Each component may vary in amplitude as a function of time and experimental manipulation. Our goal is to obtain estimates of the set of weights describing the scalp distribution of the components, <u>and</u> of the variations in amplitude as a function of time and experimental manipulations. Within the framework of our model, such estimates would provide a complete description of the ERPs under study.

Two approaches for the estimation of the set of weights for each component can be considered. One approach is based on the knowledge of the scalp distribution of an ERP component, based on previous reports or on some basic experimental paradigm. The other approach considers the sets of weights which satisfy particular criteria, such as maximization of variance explained (Principal Component Analysis approach) or optimization of the classification of sets of data (Discriminant Analysis approach). The

estimation of the amplitude of the components as a function of time and experimental manipulation can be accomplished by using the set of weights for each component as a linear filter.

To demonstrate the ability of the procedure to discriminate between overlapping components, the results obtained by applying our approach to a study of P300 in aging will be illustrated. The data suggest that the scalp distribution of the P300 peak is different for young adults and old people over a variety of tasks. Such an observation is consistent with previous reports (Pfefferbaum, Ford, Roth, and Kopell, 1980). The two groups showed both overall differences and task related differences. The analysis of the ERPs observed in each task revealed that these differences in scalp distribution cannot be entirely attributed to a generally different distribution of P300 in the two groups, or to variations in P300 amplitude, or to a combination of these two factors. Rather, they should be at least in part attributed to the presence of overlapping component(s). This is particularly evident for the subjects in the old group. A graphic representation of the variation in scalp distribution as a function of group, task and stimulus is shown in Figure 1. Such a finding suggests that subjects of the old group might use "alternative processing routes" in response to the cognitive demands of different tasks.

Another application of Vector analysis is to filter for a particular scalp distribution (Vector filter - Gratton, Coles, and Donchin, 1983). Filtering for scalp distribution improves the discrimination between signal (component) and noise (background EEG and overlapping components). The power of Vector filter is illustrated by a simulation study, in which we compared the accuracy of P300 latency estimates obtained with several techniques. Particular attention has been given to the problem of enhancing

the discrimination between signal and noise. The best discrimination, and the most accurate latency estimation, is obtained when the characteristics of both the signal and the noise are considered. Error in latency estimation for different procedures and signal-to-noise ratios are shown in Figure 2. Preparing the data with Vector filter reduces the error of latency estimation by about 20%.

In summary, this paper demonstrates how the use of the Vector Analysis approach allows us to use information about scalp distribution to distinguish between overlapping components or between components and noise. We have illustrated how the approach can be utilized to analyze between subject differences in components and to filter for a particular component.

References

Donchin, E. , Ritter, W. , and McCallum, C. (1978) Cognitive Psychophysiology : The endogenous components of the ERP. In E. Callaway, P. Tueting, and S. H. Koslow (Eds.), Event-Related Brain Potentials in Man. New York: Academic Press, 349-441.

Gratton, G., Coles, M.G.H., and Donchin, E. (1983) Filtering for scalp distribution : A new approach (Vector Filter). (Abstract) Psychophysiology, 20, 443-444.

Pfefferbaum, A., Ford, J.M., Roth, W.T., and Kopell, B.S. (1980) Age-related changes in auditory event-related potentials. Electroencephalography and Clinical Neurophysiology, 49, 266-276.

Figure Legends

Fig. 1. Graphic representation of the scalp distribution with Vector Analysis. The axes represent two orthogonal scalp distributions (schematically indicated in the figure). The ellipses represent 90% confidence intervals for each group, task, and stimulus condition.

Fig. 2. Log mean square error of latency estimation as a function of the signal-to-noise ratio for four different procedures (peak-picking at Pz, solid thin; peak-picking on Vector filtered waveforms, dashed thin; cross-correlation at Pz, solid thick; cross-correlation on Vector filtered waveforms, dashed thick), in four different conditions of component overlap.

Fig. 1

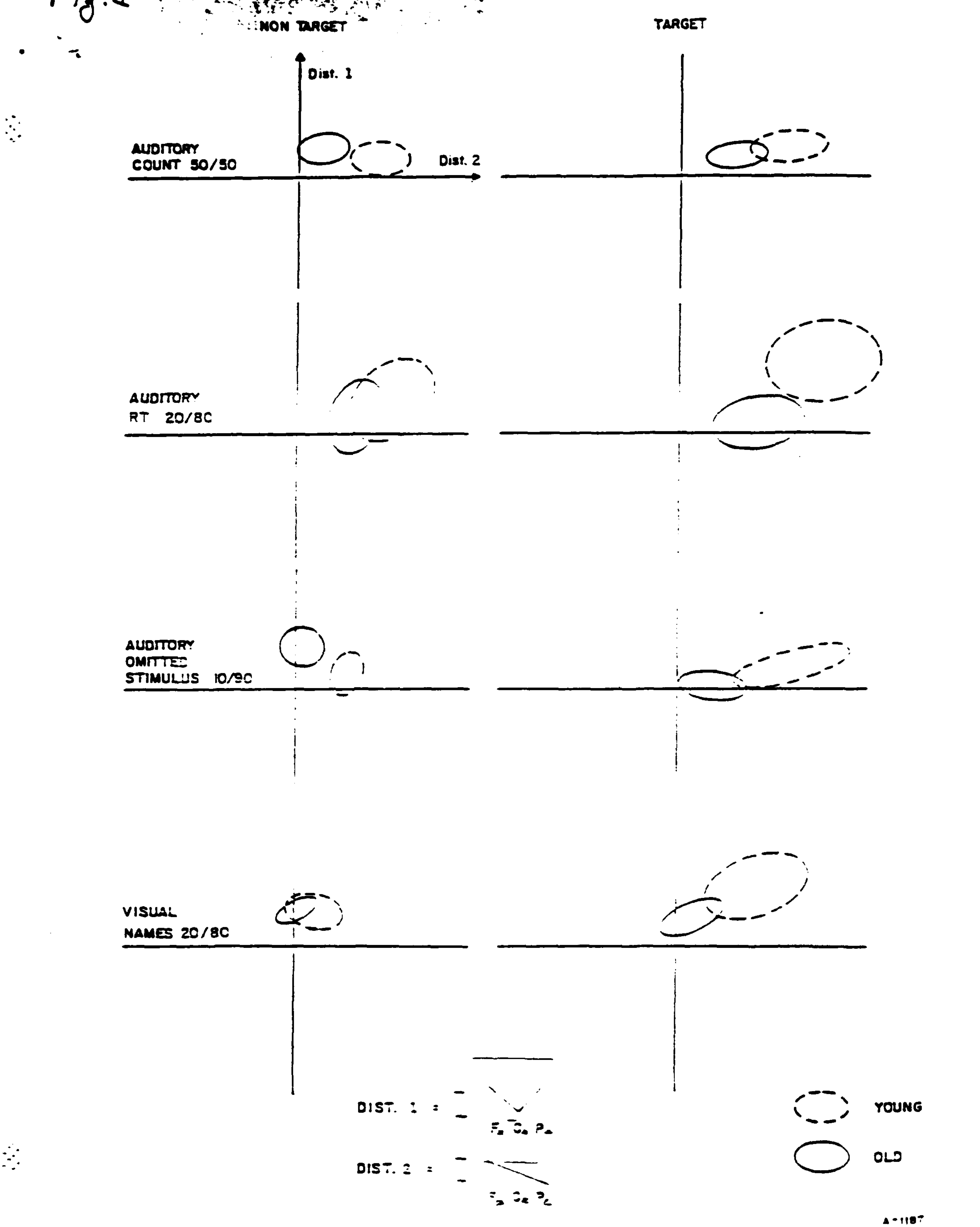
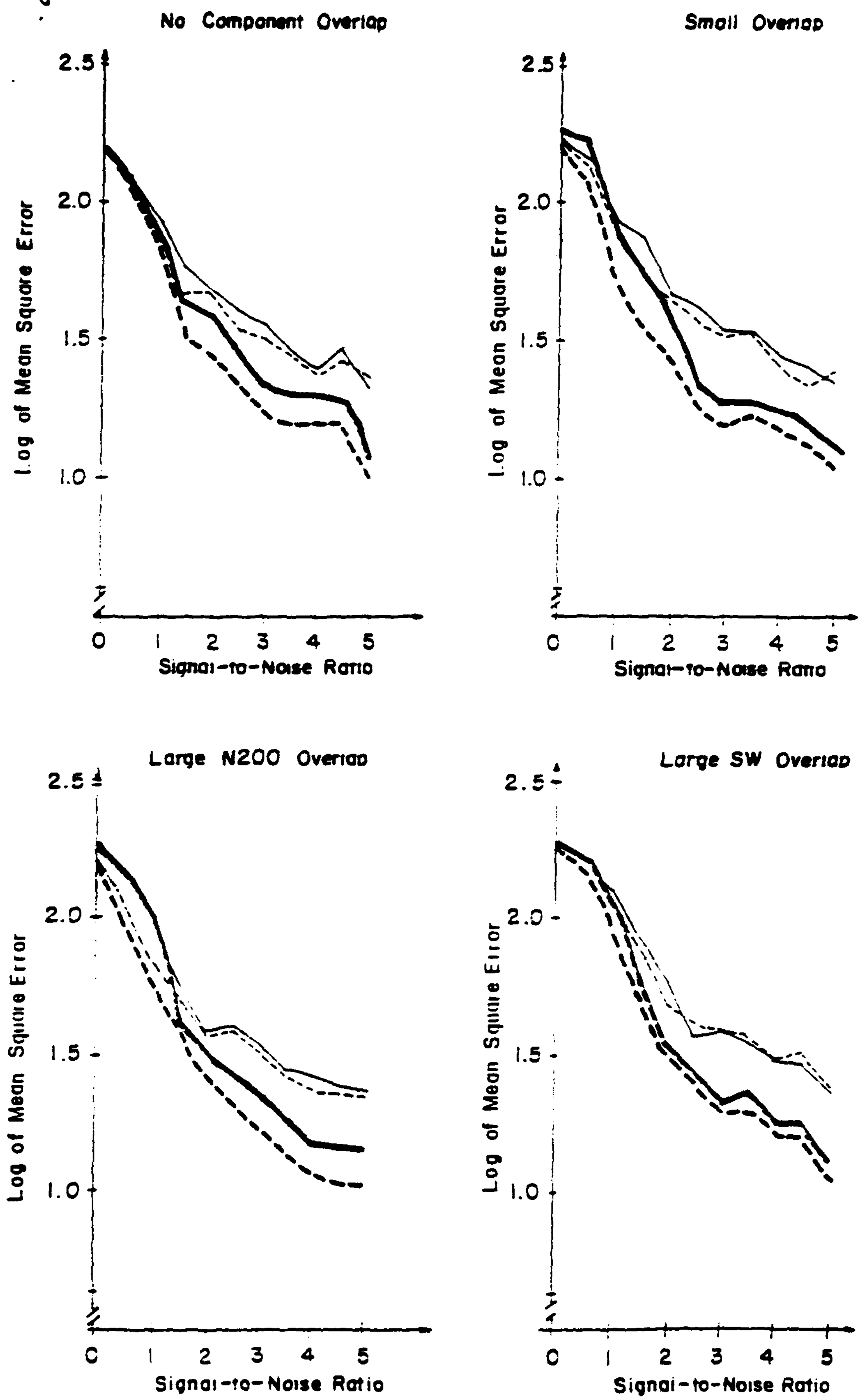

NON TARGET
TARGET
Dist. 1
Dist. 2
AUDITORY
COUNT 50/50
AUDITORY
RT 20/80
AUDITORY
OMITTED
STIMULUS 10/90
VISUAL
NAMES 20/80
DIST. 1 =
DIST. 2 =
YOUNG
OLD
A-1187

Fig. 2


No Component Overlap
Small Overlap
Large N200 Overlap
Large SW Overlap
Log of Mean Square Error
Signal-to-Noise Ratio
2.5
2.0
1.5
1.0
0
1
2
3
4
5

APPENDIX B12

P300 Amplitude and Resource Allocation

Erik Sirevaag, Arthur Kramer,

Michael G.H. Coles, & Emanuel Donchin

Manuscript in preparation, based on thesis entitled "P300 Contributions to the Analysis of Workload," submitted by E. Sirevaag in partial fulfillment for the A.M. degree, University of Illinois, 1985.

P300 CONTRIBUTIONS TO THE ANALYSIS OF WORKLOAD

Erik John Sirevaag, A.M.
Department of Psychology
University of Illinois at Urbana-Champaign, 1985

The amplitude of the P300 component of the Event-Related Potential (ERP) has proven useful in identifying the resource requirements of complex perceptual-motor tasks (Wickens et. al. 1983). In dual-task conditions, increases in primary task difficulty decrease the amplitude of secondary task P300s. Furthermore, P300s elicited by discrete primary task events increase in amplitude with increases in the difficulty of the primary task. This suggests that a reciprocal relationship between primary and secondary task P300 amplitudes should be obtained when primary and secondary task ERPs are concurrently recorded.

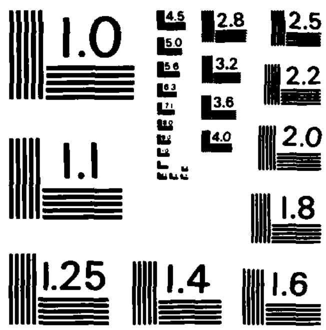
Forty subjects participated in a study designed to confirm this prediction of P300 amplitude reciprocity. Measures of subjective effort, P300 amplitude, and task performance were obtained within the context of a pursuit step tracking task performed alone and with a concurrent auditory oddball. Task difficulty was orthogonally manipulated by varying both the number of dimensions to be tracked (from one to two), and the control dynamics of the tracking task from a velocity (first order) to an acceleration (second order) system. ERPs were obtained for both secondary task tones and primary task step changes. Effort ratings and average root-mean-square (RMS) error estimates were also obtained for each tracking condition.

The data indicated that increased primary task difficulty, reflected in increased effort ratings and increased RMS error scores, was also associated with decreased secondary task P300 amplitudes and increased primary task P300 amplitudes. Since the increases in primary task P300 amplitudes were complimentary to the decrements obtained for the secondary task, the

hypothesis of reciprocity between primary and secondary task P300 amplitudes was supported across several different levels of primary task difficulty.

AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 8/9

UNCLASSIFIED 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Table of Contents

Page

Introduction...1

Definition of workload.................................1

General issues...2

Subjective effort ratings..............................4

Dual-task techniques...................................5

Multiple resource model................................6

Psychophysiological measures...........................8

Resource reciprocity..................................10

Method..14

Subjects..14

Stimuli...14

Recording system......................................15

Stimulus generation and data collection...............15

Primary and secondary task design.....................16

Procedure...16

Results...20

Strategy..20

Overt response data...................................21

Subjective effort ratings.............................23

Counting performance..................................25

Single task oddball ERP data..........................26

Primary task ERP data.................................30

Secondary task ERP data...............................36

Reciprocity confirmation..............................41

Table of contents (cont.)

Conclusions..46

References...53

Introduction

This study is concerned with the measurement of workload. Three different measures will be utilized to assess policies of processing resource allocation between two concurrently preformed tasks. These measures, the subjective rating of effort by the operator, electrophysiological concomitants of resource allocation, and measures of subject performance of an assigned task, were obtained during several step tracking conditions differing in difficulty. The workload related to various manipulations of the difficulty of the tracking task was assessed by dual task techniques (Knowles, 1963; Rolfe, 1971; and Brown, 1978).

Definition of workload Most modern theories of workload derive their theoretical basis from the concept that attention can be likened to the limited processing capacity of a general purpose computer (Moray, 1967; Kahneman, 1973; Norman and Bobrow, 1975). The allocation of this limited processing commodity to the performance of a given task is determined both by the motivation of the operator and the demand characteristics of the task. The latter can, in turn, be manipulated either by changing the nature of the task (reducing stimulus discriminability, or changing the pacing of the task, for example), or by varying the level of performance required from subjects engaged in the task.

Thus, as the difficulty of a task is increased, the workload entailed by the task is increased and more processing resources must be allocated to successfully perform the task. The concept of workload is, therefore, closely related to the allocation of processing resources. We adopt the view, presented elsewhere, that workload is "an hypothetical construct whose function is to account for decrements in performance that can not be

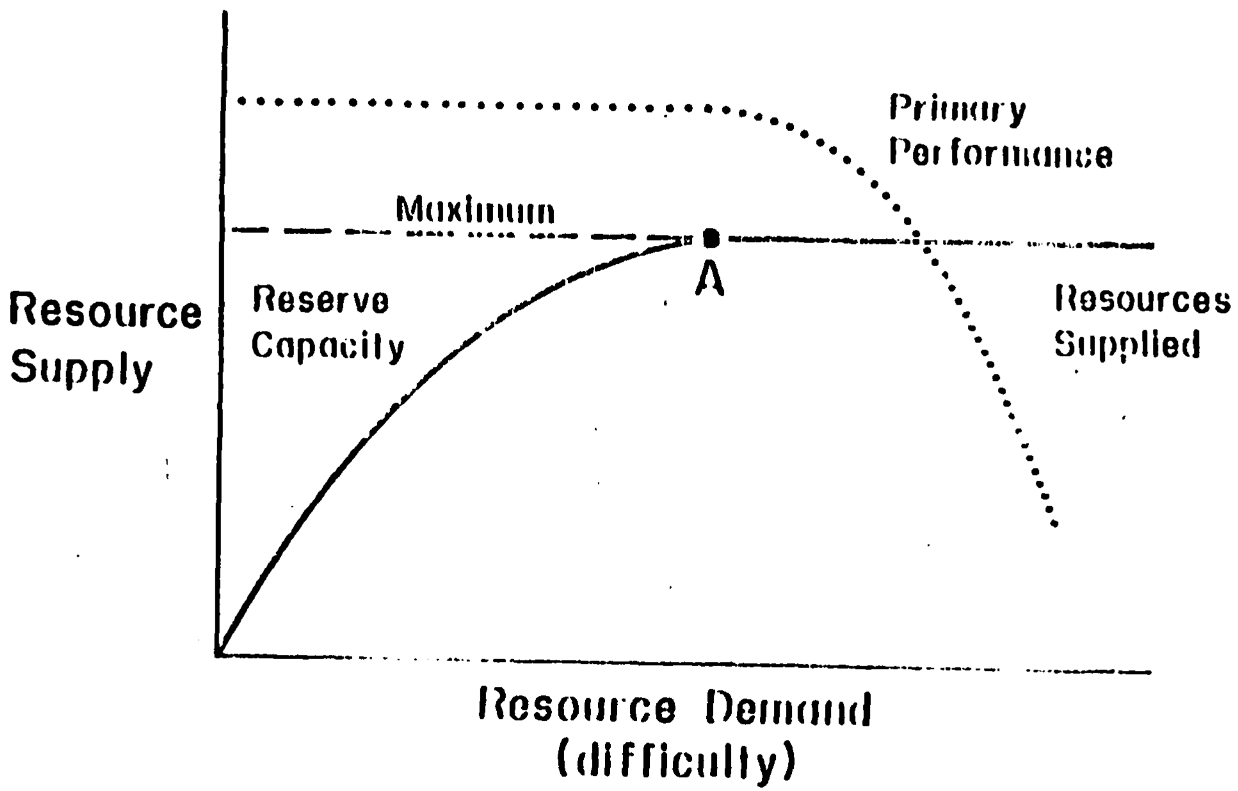
accounted for by reference to obvious limitations of the organism" (Donchin and Gopher, in press).

It is important to note, however, that although mental workload is intimately related to both task demands for resources and to performance decrements, the terms are not synonomous. The theoretical position adopted by this paper is illustrated in Fig. 1. If demands upon resources exceed the resource capacity of the operator, performance decrements will occur (the region to the right of point A). However, if fewer resources are demanded than are available, workload is related to the amount of residual capacity. Thus, in the region to the left of point A, workload is reciprocally related to reserve capacity; in the region to the right, it is reciprocally related to primary task performance (Wickens, 1984).

General Issues The assessment of workload is one of the critical issues in Engineering Psychology. Wickens (1981) has proposed that measures of mental workload should satisfy the following five criteria: (a) "sensitivity", implies that the measure should be sensitive to graded changes in task difficulty; (b) "diagnosticity", is satisfied if, in addition to sensing variations in workload, the measure indicates the source of such variation; (c) "selectivity", requires that workload measures be sensitive only to differences in capacity demand and should not reflect changes such as physical load or emotional stress that are unrelated to mental workload (according to some theories); (d) "unobtrusiveness", states an ideal workload index should not interfere with performance of the task whose workload is to be assessed; and (e) "bandwidth and reliability", indicate that measures assessing workload in a time-varying environment should be available sufficiently rapidly so that transient changes can be reliably estimated.

Figure 1

The Supply and Demand Model of Workload



This figure illustrates the relationship between the allocation of resources and task performance. Workload is reciprocally related to reserve capacity in the region to the left of point A; and is reciprocally related to performance in the region to the right of point A (from Wickens, 1984)

Subjective effort ratings Measures of single task performance are not suitable workload metrics for two reasons: they cannot be generalized across tasks requiring different performance measures; and increased task difficulty is often not reflected in single task performance decrements (presumably due to the increased allocation of processing resources). Therefore, numerous alternative techniques have been proposed for workload assessment (for reviews see Wickens, 1984; Gopher and Donchin, in press; Williges and Wierwille, 1979). Sheridan (1980) has argued that subjective ratings, or individual reports based on subjective experience of the cognitive effort entailed by a particular task, come the closest to tapping the essence of mental workload. A variety of studies (reviewed by Borg, 1978; Ellis, 1978; and Moray, 1982) have attempted to assess the utility of subjective measures of mental workload. The problem with subjective ratings is that not only is it difficult to assess the accuracy and reliability of such reports, it is often difficult to distinguish the dimensions of the tasks upon which the reports are based, making diagnosis difficult if not impossible.

Many of these difficulties can be overcome, however, when multidimensional scaling techniques are employed (Derrick,1981). Thus, the work of Derrick indicates that although great care must be taken whenever subjective ratings are to be related to workload, the dimensions underlying such reports can be reconstructed if appropriate scaling techniques are employed. However, recent evidence (Yeh, Y. and Wickens, C., 1984) indicates that under certain conditions subjective ratings are overly sensitive to manipulations of task difficulty and under sensitive in other situations.

Dual-task techniques An alternative approach stresses the advantages of dual-task techniques of workload assessment (Brown, 1978; Knowles, 1963; Rolfe, 1971). If resources represent a limited commodity allocated according to task demands and performance requirements, then two tasks performed concurrently must compete for these limited resources. Furthermore, if subjects are instructed to consider one of the concurrent tasks as the primary task, fewer resources will be available for the secondary task. Therefore, increases in primary task difficulty should entail greater resource allocation to the primary task accompanied by secondary task performance decrements due to this drain on resources.

A number of studies have both refined and validated the dual task paradigm as a method for workload assessment under a variety of different conditions (Kahneman, 1973; Navon and Gopher, 1979; Norman and Bobrow, 1975; Schneider and Fisk, 1981; Sperling and Melchner, 1978). However, several other studies indicate that the model of processing resources as a single pool of undifferentiated capacity is not completely satisfactory. In an experiment conducted by Wickens (1976), greater performance decrements within a tracking task were produced by the introduction of a concurrent task requiring a pure response (maintaining constant pressure on a stick), than by a concurrent signal detection task, even though subjects rated the latter as the more difficult condition. Thus, interference effects between two tasks can not always be specified by their relative difficulty, as the undifferentiated capacity model would predict.

Furthermore, increases in the difficulty of one task are not always associated with decreased performance in the concurrent task. For example, in a study by North (1977), subjects performed a tracking task and a discrete response task involving mental operations of varying complexity

upon visually displayed digits. Although single task performance of the various digit tasks indicated that they entailed differential workloads, all of the conditions disrupted performance of the tracking task to the same degree. This problem of difficulty insensitivity also indicates the inadequacy of a theory involving a single pool of processing resources.

Finally, two tasks that are clearly attention demanding can be time-shared perfectly. Allport, Antonis, and Reynolds (1972), for example, asked skilled pianists to shadow verbal material while sightreading music. Both tasks were performed at single task performance levels. In addition, Shaffer (1975) has demonstrated that skilled typists can display perfect time sharing when transcribing written material while performing a concurrent shadowing task. Once again, perfect time-sharing is easily explained if independent resource pools are postulated, but is difficult to explain in terms of a single resource pool of undifferentiated capacity.

Multiple resource model Wickens (1983) has offered an alternative model to that of Kahneman. This model postulates multiple pools of resources. In Wickens' model, displayed in Fig.2, three largely independent sets of resource pools coexist with undifferentiated capacity. These separate resource pools can be defined in terms of three dichotomus dimensions: one related to different stages of processing (perceptual vs. response selection processes, for example); one related to the encoding structure of the task (spatial vs. verbal); and one related to input/output modality (auditory vs. visual).

Although the multiple resource model does not predict how the pools may interact, a large proportion of the variance between the findings of various dual-task studies can be explained if more than one pool of resources is assumed to exist. However, even by utilizing a multiple resource theory

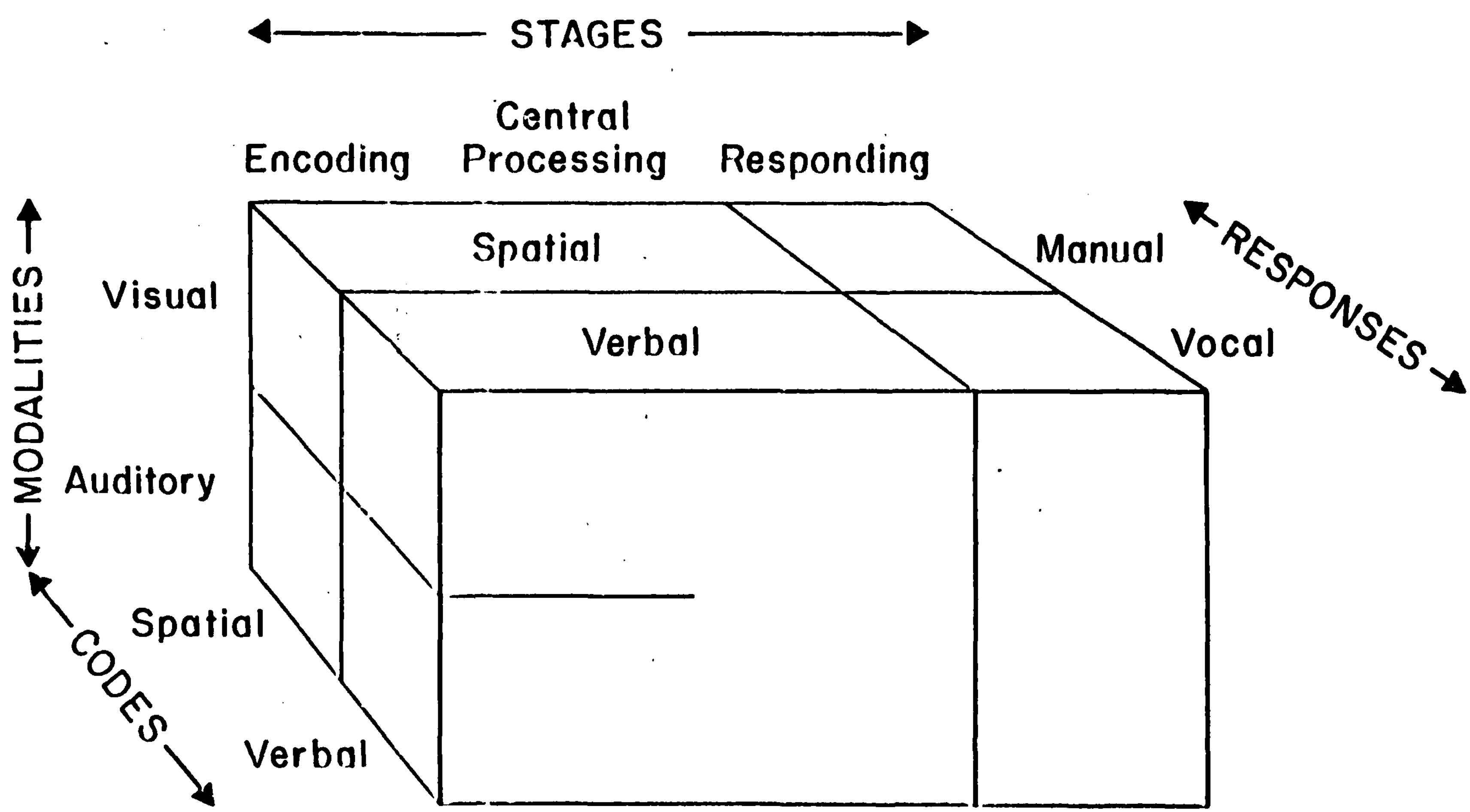


Figure 2

The Multiple Resource Model

This figure depicts Wickens' multiple recource model of processing capacity (from Wickens, 1984).

for secondary task design and analysis, there are several practical problems in implementing a dual-task study (Brown, 1978; Ogden, Levine, and Elsner 1979). The most immediate problem is that in general it is exceedingly difficult to design secondary tasks so that the responses required do not impede performance of the primary task. Although such difficulties are not insurmountable, when studying the perceptual-central processing demands of a primary task, it is preferable to minimize secondary task response conflict.

Psychophysiological measures As a consequence, several investigators suggested the use of physiological measures as indices of workload which do not require the subject to respond overtly to a secondary task (Berlyne, 1960; Howitt, 1968; Roscoe, 1978; Wierwille, 1979). In the main, these measures were to be used as presumably direct measures of arousal. However, it is also possible to use a psychophysiological measure, the Event-Related-Brain-Potential (ERP) within the framework of the secondary task paradigm. It is on this approach that the present study is founded.

The ERP is obtained by averaging the digitized values of electroencephalographic (EEG) activity time-locked to an event (Donchin, Ritter, and McCallum, 1978). By averaging over several repetitions of the event, background activity unrelated to the processing of the event diminishes while the time-locked activity is enhanced.

Previous dual-task studies reviewed by Donchin, Kramer, and Wickens in 1982, provided evidence that attributes of one component of the ERP, the P300, vary as a function of workload. The P300 is a positive voltage deflection maximal over parietal scalp with a minimal latency of 300 msec. (Sutton et. al. 1965). These studies assume, of course, a limited capacity resource theory: an operator has pools of limited resources at his disposal during the performance of a task; and more difficult tasks require more

resources if performance levels are to be maintained. Therefore, increases in difficulty should be reflected in secondary task decrements when the two tasks compete for resources from common pools. Tasks in which the P300 component of the ERP can serve to measure secondary task performance have an advantage as there are no overt response requirements in such tasks.

Secondary tasks employing the "oddball" paradigm and measures of the event-related potential have been particularly successful. In a typical oddball task, subjects are asked to discriminate between two stimuli differing along some dimension (for example, two tones differing in pitch). One of the stimuli is designated as the target, the other as the non-target. occurrences of non-targets are to be ignored; while occurrences of the targets are to be counted silently and reported at the end of the block of trials.

Given that the amplitude of the P300 is proportional to the extent to which a subject utilizes the information provided by a stimulus (Johnson and Donchin, 1978; Johnson and Donchin, 1982; Duncan-Johnson and Donchin, 1977; Donchin, Kubovy, Kutas, Johnson, and Herning, 1973), and that the latency of the P300 has been shown to be sensitive to stimulus evaluation processes and relatively insensitive to response selection processes (Kutas, McCarthy, and Donchin, 1977; see also Donchin and Isreal, 1979), it seems reasonable to suppose that this component may serve as an index of the relative relevance of the oddball task.

Thus, because the amplitude of the P300 is proportional to the subjective probability of stimuli on a trial-by-trial basis, an essential component driving single trial P300 amplitude must be the extent to which a subject is actually allocating resources towards the performance of the task. Furthermore, because aspects of the P300 have been shown to be

selectively sensitive to manipulations of stimulus evaluation, the P300 should provide a metric of resource competition that is limited to stage-specific (ie central/perceptual) pools of processing resources. Further support for these assertions comes from the fact that oddball stimuli fail to produce P300s if subjects are instructed to ignore the tones (Donchin and Cohen 1967; Duncan-Johnson and Donchin, 1977; Ford, Roth, and Koppell, 1976). Thus, reductions in P300 amplitude to attended secondary task tones related to increases in primary task difficulty are presumed to reflect increased resource allocation to the primary task.

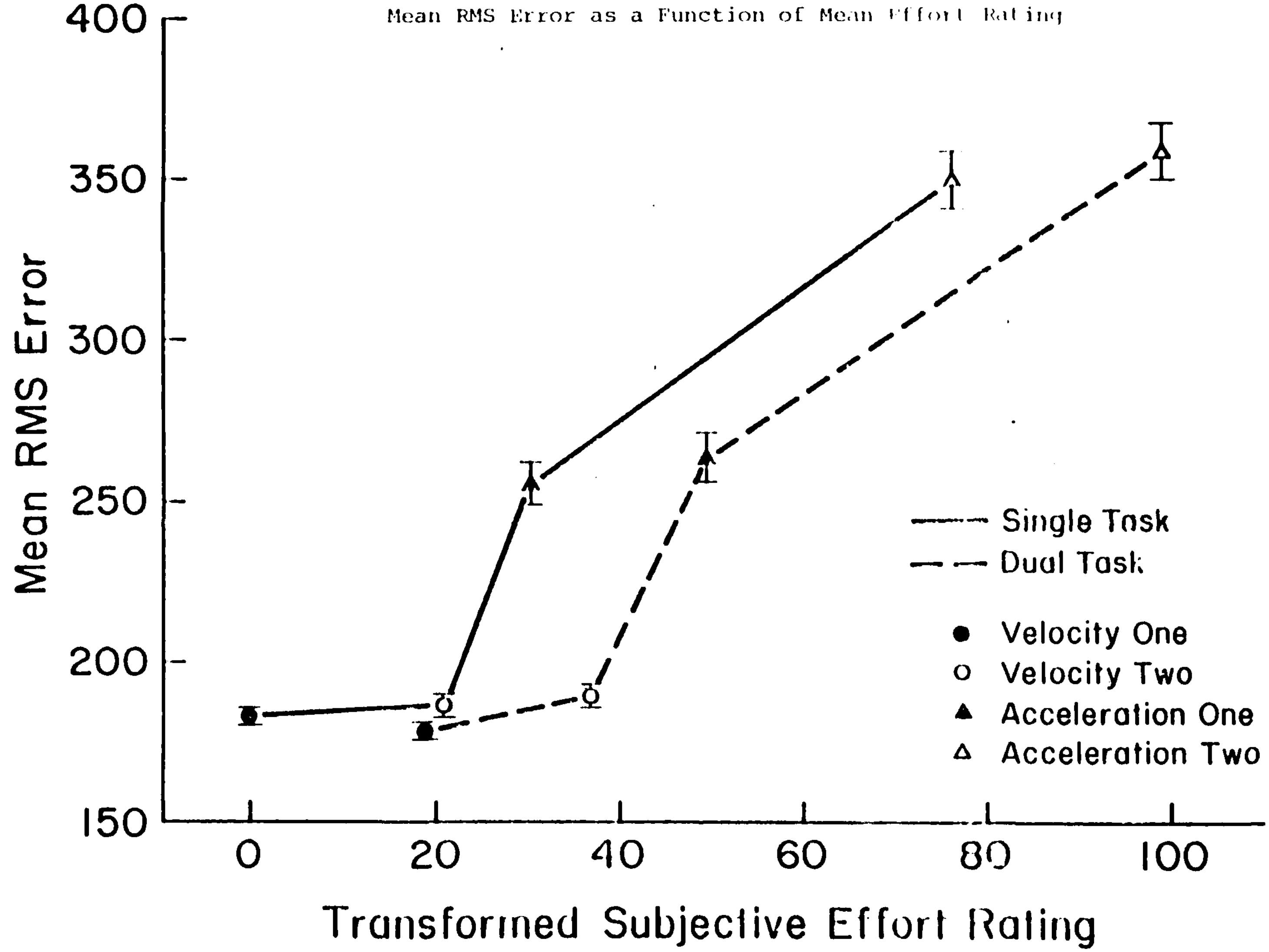
Resource reciprocity The utility of P300 as a metric of mental workload has been examined, and a considerable body of evidence has accumulated indicating that P300 amplitude is sensitive to competition for limited resources in the perceptual domain (Isreal, Wickens, Chesney, and Donchin 1980; Isreal, Chesney, Wickens, and Donchin, 1980; Kramer, Wickens, and Donchin 1983). This implies that increased allocation of resources to primary tasks should be reflected not only in secondary task decrements, but also in increased primary task P300 amplitudes.

Thus, in dual-task studies in which ERPs can be recorded in response to discrete primary and secondary task events, there should be a reciprocal relationship between primary and secondary task P300 amplitudes. The term "resource reciprocity" has been suggested to describe this relationship. As additional perceptual resources are allocated to the primary task, secondary task P300s should decline and primary task P300s should increase in amplitude if the P300 is indeed a metric of resource allocation.

In a pursuit step tracking study, (Wickens, Kramer, Vanasse, and Donchin, 1983) this concept of P300 amplitude reciprocity was explicitly tested. A reciprocal relationship between primary and secondary task P300

Figure 4

Mean RMS Error as a Function of Mean Effort Rating



The average RMS tracking error is plotted as a function of the average transformed effort rating reported by the subjects for a given tracking condition.

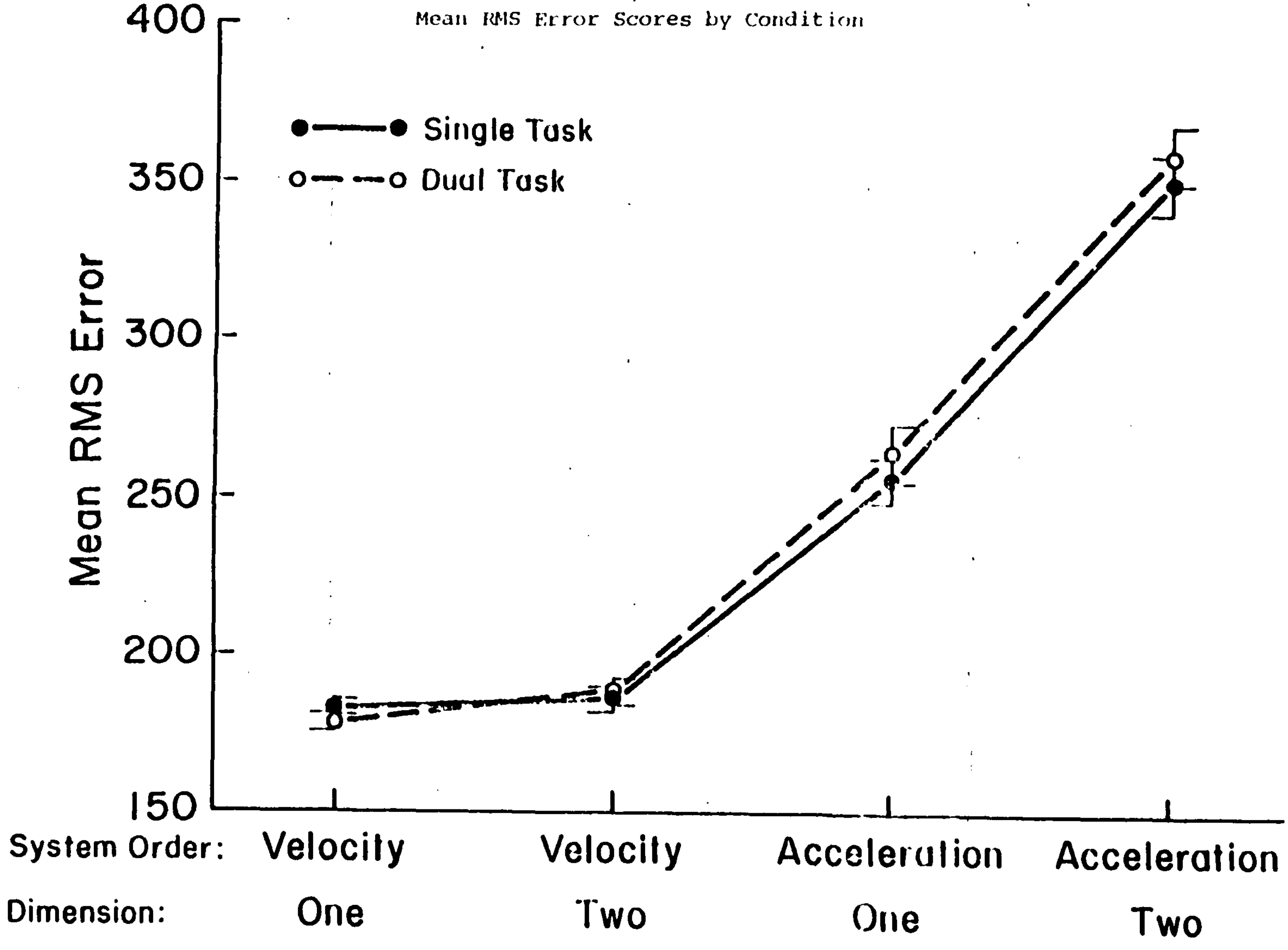
dimensionality was significant only when the tracking task required acceleration control ($p<.01$).

Subjective effort ratings There is substantial agreement between the subjective ratings of workload and the RMS error data. Subjects consistently perceived dual tasks as requiring greater effort than single tasks [$F(1,39)=158.27, p<.01$]. Second order systems required more effort than first order tracking tasks [$F(1,39)=550.02, p<.01$]; and tracking in one dimension was seen as easier than tracking in two [$F(1,39)=515.78, p<.01$].

Although Tukey tests of specific pairwise comparisons indicate that subjects reported higher ratings for the order manipulation in both dimensions ($p<.01$), as in the RMS data, the magnitude of the order effect was much larger in two dimensional tracking tasks than in one dimensional tracking blocks. This produced a significant dimension x order interaction [$F(1,39)=48.58, p<.01$].

The relationship between the average RMS error scores and the effort ratings is summarized in Fig. 4. Note that although there is general agreement between the effort ratings and the performance data, subjects' perceptions of the difficulty of the task are not necessarily consistent with their ability to perform the task: In fact, subjects consistently reported that dual tasks were more effortful than single tasks, yet they tracked in both with about equal accuracy. These findings are consonant with those of Yeh et. al. (1984) which indicate that subjective ratings tend to be overly sensitive to the imposition of a secondary task. Furthermore, in the present experiment, subjects claimed an increase in effort as the number of dimensions increased, regardless of system order; however, dimensionality affected their performance only in second order systems. This indicates an over-sensitivity to a different manipulation of task

Figure 3
Mean RMS Error Scores by Condition



This figure displays the mean RMS tracking error for each of the tracking conditions averaged across all subjects.

presentation will be evaluated to determine whether increased primary task P300 amplitudes were associated with decreased secondary task P300s. Primary and secondary task P300 amplitude reciprocity will be evaluated both across and within individual subjects.

Overt response data The average root mean square (RMS) error for a given block is proportional to the average distance between the subject controlled cursor and the target square. Low RMS error scores, therefore, reflect increased tracking accuracy. The RMS error data were evaluated to assess the quality of subject performance. To determine if RMS error varied significantly with the number of dimensions tracked, and with the order of the system, we compared the RMS error averaged within the blocks of trials. The relevent means are presented in Fig. 3. In all subsequent figures, deviations from the means represent the standard error of estimate.

Note that subjects' ability to track did not decline when the secondary task was imposed. The increase in RMS error associated with dual tasks, albeit statistically significant [$F=(1,39)=4.94, P<.05$], was less than 2% of the RMS error during the single task (and when the mean for each subject was divided by that subject's standard deviation the difference disappears altogether). Tracking accuracy declined both as dimensionality increased [$F(1,39)=243.81, p<.01$], and as the control order was increased from a velocity to an acceleration system [$F(1,39)=408.52, p<.01$]. The effect of system order was consistently larger than the effect of dimensionality. It is noteworthy that dimensionality affects the accuracy of tracking largely for the second order systems as can be seen from the interaction between order and dimensionality [$F(1,39)=190.98, p<.01$]. Tukey tests performed on specific pairwise comparisons confirm that while order significantly affected performance in both one and two dimensions, the effect of

Results

Strategy The subjective effort ratings and primary and secondary task overt performance data will first be examined to assess the extent to which the variations in dimensionality and in system order indeed modulated the performance of the primary task, as well as to determine whether subjects did in fact protect the level of performance on the primary task even as the task demands increased. This is a critical observation if the dual-task methodology is to be applied. Secondary task performance decrements cannot be properly interpreted if subjects did not, in fact, maintain their performance of the primary task.

The ERPs elicited during the single task visual and auditory oddball paradigms will be described next. We will show that the stimuli employed by the primary and secondary tasks elicited P300s within the conventional latency range, and with the conventional morphology, and sensitivity to task demands. Having established this we will compare the ERPs elicited in the single and dual primary tasks. If subjects did in fact protect their performance of the tracking task, P300 amplitudes should not be drastically reduced in the dual task conditions from the amplitude levels established during the track alone conditions.

With these observations established, the dual task waveforms can be analyzed to assess the effects of increased primary task difficulty upon the amplitude of P300s associated with the step changes of the primary task. Recall that we predicted that increased primary task P300 amplitudes will be associated with increases in the difficulty of the tracking task. Confirming this prediction is consistent with the existence of resource reciprocity. Finally, the P300s associated with secondary task tone

In addition, on every single and dual-task block, data concerning tracking accuracy was collected for every trial by sampling the root-mean-square (RMS) error between the subject controlled cursor and the target square every 50 msec. The final measure of performance accuracy, counting task error rates, was collected during all dual-task conditions.

The ERPs were recorded during all series. Parietal, central, and frontal electrode outputs were digitized and recorded to the step changes during step count, single task tracking, and dual-task tracking conditions. ERPs to the tones were recorded during the single task count oddball as well as to the oddball tones counted while subjects were in the dual-task tracking conditions.

The data were submitted to a range correction algorithm when such a procedure was required to either facilitate the comparison between two different metrics or to correct for large between subject variability. For example, subjects varied greatly concerning the range of effort ratings they employed to describe the difficulty of the various conditions. Comparisons across subjects would not have been possible without first transforming the data. The following formula was used to accomplish this transformation:

$$X(T) = 100 * \frac{X(I) - X(MIN)}{X(RNG)}$$

where, X(T)= transformed score; X(I)= single block score;
X(MIN)= minimum score for a given subject;
X(RNG)= range of scores for a given subject.

As a result of this transformation a subjects minimum score recieved a value of 0 and his maximum score a value of 100.

presentation was constrained so that the recording epochs of the tones and the step changes did not overlap.

The blocks were presented in the fixed order displayed in Table 1.

Table 1
Order of task presentation

	velocity in one dimension	acceleration in one dimension	velocity in two dimensions	acceleration in two dimensions
single	1			
dual	2			
single			3	
dual			4	
single		5		
dual		6		
single				7
dual				8

Note that single task blocks always preceded dual-task blocks, and easier tracking conditions preceded difficult tracking conditions. This order was designed to hold constant the order in which subjects experienced the various levels of difficulty. Task difficulty and presentation order were, therefore, deliberately confounded. The order displayed in Table 1 was chosen so that any learning due to practice effects should improve performance during the more difficult conditions, rather than enhance performance decrements due to increased primary task difficulty.

Effort ratings were taken after every block. Subjects were asked to "provide a numerical estimate of how difficult the preceding block was to perform". To provide a common anchoring point, the subjects were instructed to rate the effort entailed by the task on a scale where the difficulty of a previously administered symbol-digits modality task (Smith, 1973) was given a value of 100.

One of the counting tasks was auditory involving the same tones used in the dual-task oddball. In this task, targets and nontargets were equiprobable (p=.50). This condition allowed for an assessment under single task conditions of the extent to which standard P300s were associated with the stimuli employed in the dual task oddball conditions. The other counting task was in the visual modality and involved step changes identical to those in the primary step tracking task. Half the subjects were instructed to count movements of the computer-controlled square to the left, the other half counted step changes to the right. This condition was included to demonstrate that the primary task stimuli could produce P300s under standard oddball conditions requiring no overt response.

Subjects were then given three practice single task tracking blocks. These consisted of one block of velocity tracking in two dimensions, and two blocks of acceleration tracking (one in one dimension and one in two dimensions). After completing the practice blocks, subjects were instructed to consider the tracking task as primary. They were told that while counting accuracy was important, their primary goal was to perform the tracking task as well as possible. In addition, the following bonuses were made available: fifteen cents every time the tone counts reported were within one of the correct count; and one dollar if the experimenter determined that the subject was at least attempting to perform the tracking task for the entire duration of each tracking block.

The eight experimental blocks were then run. Each single task block contained, on the average, 60 step change trials presented at an average inter-stimulus interval (ISI) of 6 seconds. Each dual-task block contained an average of 60 tone trials in addition to the 60 step changes. Tone

EOG to the EEG waveforms were evaluated and eliminated off line by submitting the data to an eye-movement correction algorithm developed by Gratton et. al. (1983).

Primary and secondary task design In single task conditions, two squares were visually presented on a screen in front of the subjects. Their task was to superimpose the square under their control (cursor) upon the square under computer control (target) as rapidly as possible.

System order (first order velocity vs. second order acceleration); and dimensionality (tracking horizontally vs. tracking both horizontally and vertically) were orthogonally manipulated creating four single task conditions. In the dual-task conditions, subjects were instructed to perform an oddball tone counting task concurrently with the tracking task. Therefore, there were eight (four single and four dual) blocks of tracking trials administered to each subject.

Procedure Each of the forty subjects participated in all of the experimental conditions. Each condition lasted approximately 15 minutes and was followed by a short (2 min.) break. Following electrode placement, subjects were instructed that they were about to participate in a study to assess the effects of increased task difficulty under both single and dual-task conditions.

Before receiving practice on the step tracking task, all subjects performed three oddball tasks in order to familiarize themselves with the stimuli, as well as to demonstrate that the data we were recording were similar to results obtained previously (Sutton, 1965). In all of the counting tasks, the stimuli and the rates of stimulus presentation were identical to those under dual-task conditions.

system, "A" equalled zero. This was, therefore, a pure first order system. In such a system the movements of the subject's cursor were directly proportional to the position of the joystick. To bring the cursor to a stop, the subject had only to allow the joystick to rest in the center position with zero deflection. For acceleration conditions, A = 95, so the system order was actually a linear combination of first and second order systems. Under these circumstances, a joystick deflection in the opposite direction of the cursor's movement was necessary to bring the cursor to rest.

<u>Recording system</u> Electroencephalographic (EEG) activity was recorded from three midline sites (Fz, Cz, and Pz according to the 10-20 system: Jasper, 1958) referenced to linked mastoids. Two ground electrodes were attached to the forehead. The scalp and mastoid electrodes were all Burden Ag-AgCl electrodes affixed with collodion. The vertical electro-oculogram (EOG) was recorded from Beckman Biopotential electrodes affixed with adhesive collars above and below the subject's right eye. Beckman electrodes were also used for the grounds. All electrode impedances were below 10 kohms/cm.

The EEG and EOG were amplified by Van Gogh model 50000 amplifiers with a 10 sec. time constant and an upper half amplitude of 35 Hz, 3db/octave rolloff. The recording epoch for both the EEG and EOG was 1280 msec. and began 100 msec. prior to either the primary task step changes or the secondary task tones. The data channels were digitized every 10 msec. and were also filtered off-line (-3db at 6.29 hz.) prior to further analysis.

<u>Stimulus generation and data collection</u> Presentation of the stimuli and collection of the data were under the control of a PDP 11/40 computer (see Donchin and Heffley, 1975). On line monitoring of both average and single trial EEG and EOG was accomplished by a GT-44 display. Contributions of the

Method

<u>Subjects</u> Forty dextral males between the ages of 18 and 25 were paid for their participation in this study. None of the subjects had any previous experience with the step tracking task. All subjects had normal hearing and normal or corrected to normal vision.

<u>Stimuli</u> Auditory stimuli were presented binaurally through TDH-39 headphones. Presentation of a low pitched tone (1200 hz.) alternated randomly with a high pitched tone(1400 hz.). The duration of both tones was 60 msec. (including a 10 msec. rise/fall time) and the two tones were equiprobable.

The target and cursor were both square and were generated and displayed by an IMLAC graphics computer on a screen located directly in front of the subject. Movements of the target square were under the control of the computer. These movements consisted of discrete jumps to random positions on the screen. Such jumps could occur in either the horizontal or both the horizontal and the vertical dimensions depending on the requirements of a given condition. Jumps were constrained so that there were equal numbers of jumps in all directions in a given block.

Subjects controlled the position of the cursor by manipulating a joystick with their right hand. The dynamics of the system response to movements of the joystick were determined by the following equation:

$$X(T)=[(1-A)\ U(T)\ dt]+[(A)\ \ U(T)\ dt]$$

where U=stick position; T=time and A=difficulty level.

This equation altered the number of time integrations between the joystick output and the movements of the subject's cursor. Thus, for the velocity

task difficulty to be evaluated in terms of both subjective effort ratings and primary task performance.

increased primary task P300s related to decreased secondary task P300s. It is equally important, however, to examine the relationship between primary and secondary task P300 amplitude under conditions where the manipulation of task difficulty has no effect upon secondary task P300 amplitude. Thus, when the secondary task P300 data indicate that difficulty did not vary between two conditions within the perceptual-central domain, it is vital to demonstrate a corresponding lack of effect upon primary task P300 amplitude.

Previous research has indicated that while P300 amplitude is sensitive to increases in the system order of a tracking task (Wickens et. al. 1983), manipulations of the number of dimensions in which a subject is required to track produce no changes in secondary task P300 amplitude (Wickens, Isreal, and Donchin, 1977). Therefore, an orthogonal manipulation of dimensionality and system order should provide conditions in which the presence of resource reciprocity can be confirmed under different conditions with varying degrees of primary and secondary task resource competition.

A step tracking task was developed in which subjects were run through four conditions (2 system orders x 2 dimensions) within the context of both single and dual-task instructions (ie. the presence or absence of a concurrent auditory oddball task). In addition to clarifying the stage-specific processing requirements of manual control systems of differing levels of both dimensionality and system order, such a design provides a unique opportunity to examine whether the reciprocity of P300 amplitude can be demonstrated in conditions involving orthogonal combinations of dependent variables and concurrently recorded primary and secondary task ERPs. Because subjective reports of task difficulty were obtained after every block, this design also allows the efficacy of the manipulations of primary

amplitude with respect to manipulations of the system order of the primary task was found. Previous studies have indicated that the manipulation of system order places demands on both response related (North, 1977; Trumbo, Noble, and Swink, 1967; Vidulich and Wickens, 1981) as well as perceptual resources (Wickens, Derrick, Micalizzi, and Beringer, 1980).

In this experiment, subjects were required to monitor discrete movements along a horizontal line of a visually displayed target square. The task was to manipulate a joystick to superimpose a cursor square upon the target. ERPs time-locked to the step changes of the primary task were digitized and recorded in one condition; while those elicited by the tones counted during the secondary task were recorded in a separate condition. System order was varied by manipulating the number of time integrations between the joystick output and the movements of the cursor on the screen. The data confirmed that P300s associated with the step changes increased in amplitude with increased primary task difficulty; while secondary task P300 amplitude decreased in a complimentary manner.

The evidence for amplitude reciprocity obtained by Wicken's et. al. in the above study would be stronger if the ERPs elicited by the visual primary task and the auditory secondary task had been recorded within the context of a single experimental condition. In other words, the case for resource reciprocity can be made more strongly if a reciprocal relationship between concurrently recorded primary and secondary task ERPs is found. One of the main goals of the present study is to determine if such a relationship does in fact exist when both step changes and tones are recorded concurrently within a dual-task paradigm.

The study by Wickens et. al. (1983) cited above provides crucial evidence for the reciprocity of processing resources by demonstrating

difficulty. These small dissociations between the effort ratings and performance data are not particularly surprising. In fact, given the large body of evidence indicating that these metrics tap different sources of variance, one should be more surprised when they are in complete agreement than when they differ.

In summary, the effort ratings and the rms error data indicate that the manipulations of control order and dimensionality successfully produced a range of tracking conditions suitable for the analysis of P300 amplitude reciprocity under varying levels of primary and secondary task competition for processing resources. Furthermore, the RMS data confirm that subjects protected their performance of the primary task, for although there was a significant increase in RMS scores due to the imposition of the secondary task, this increase was trivial.

Counting performance It will be recalled that the secondary task required subjects to count one of two equiprobable tones differing in pitch while they were performing the primary step tracking task. For each block, then, the accuracy with which the subject counted could be scored. These scores were also submitted to a two-way (Dimension by Order) ANOVA to determine the extent to which counting was affected by system order and dimensionality. This analysis is crucial if one is to be sure that P300 amplitude variability associated with the secondary task is due to variations in primary task difficulty and not due to unsatisfactory counting performance. The mean number of counting errors per condition are presented in Table 2. Only the change of system order (from a velocity to an acceleration system) affected counting accuracy [$F(1,39)=8.88, p < .01$]. However, even in the most difficult condition (two dimensions with acceleration control)

subjects averaged less than two errors per block. The data suggest that subjects counted the secondary task stimuli to the best of their ability.

Table 2
Mean number of counting errors per condition

velocity in one dimension	acceleration in one dimension	velocity in two dimensions	acceleration in two dimensions
0.85	1.73	1.10	1.58

The ERP data will be examined next to determine whether the increased workload demands indicated by the performance data and the subjective effort ratings can be related to increased primary task P300 amplitude and decreased secondary task P300 amplitude.

Single Task Oddball ERP Data As outlined above, subjects were exposed to two oddball tasks before they performed any step tracking. One of the oddballs was auditory and utilized the same equiprobable tones that were used in the secondary task in the dual task conditions. The other oddball series used visual stimuli. In this task subjects merely counted step changes to either the left or the right. These conditions were included to show that the stimuli employed in the tasks could be associated with typical ERP components. The waveforms from which the ERP components were extracted represent single trial averages (based on thirty to sixty trials) from each experimental condition.

Fig. 5 shows the ERP, averaged over all subjects, elicited by the single task tones in the count only condition. Visual inspection of the waveforms confirms that a parietally maximum positive deflection is present in the latency range appropriate for the P300. Although both the targets and the nontargets display large P300s, P300 amplitude is slightly larger

for the targets than the nontargets. The small size of this target effect is common when the two tones are equiprobable.

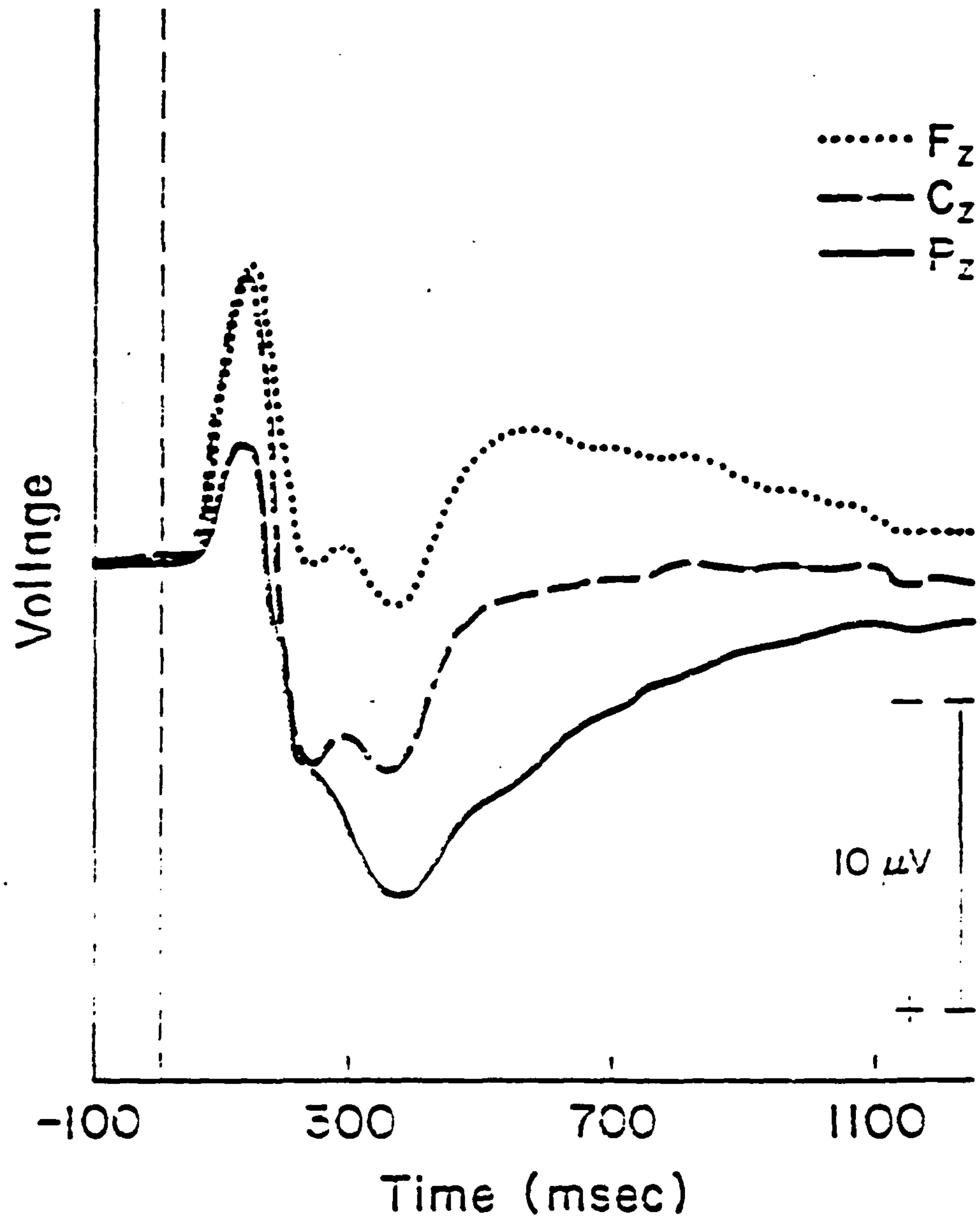
A separate Principal Components Analysis (PCA) was performed on the data from the single task auditory oddball. A matrix containing 240 trials (40 subjects x 2 Target levels x 3 electrodes) was submitted to a PCA in which five of the extracted components were Varimax rotated.

Donchin et. al. (1978, 1982) has suggested that ERP components be identified according to three criteria: their latency relative to a stimulus or a response; their amplitude distribution across different electrode sites; and their sensitivity to task manipulations. On the basis of these criteria, one component could be identified as P300 in in the single task auditory oddball condition. Subsequent repeated measures ANOVAs performed on these component scores indicated that this component was differentially distributed across electrode locations [$F(1/39)=99.75, p<.01$] with maximum positivity occurring at the parietal site. P300 amplitude was significantly larger for targets [$F(2/78)=12.86, p<.01$] at Pz and the component identified occurred in the appropriate latency range (300 to 500 msec.). Thus, the particular auditory oddball employed in this experiment produced P300s with the traditional latency, scalp distribution and sensitivity to task manipulations.

For the analysis of the single task step count visual oddball, a 240 trial data matrix constituted identically to the auditory oddball matrix was submitted to a PCA in which four of the components extracted were Varimax rotated. The grand average ERP associated with step changes is presented in Fig. 6.

The Anova performed on the component scores associated with the visual oddball indicates that one component was maximal over parietal scalp

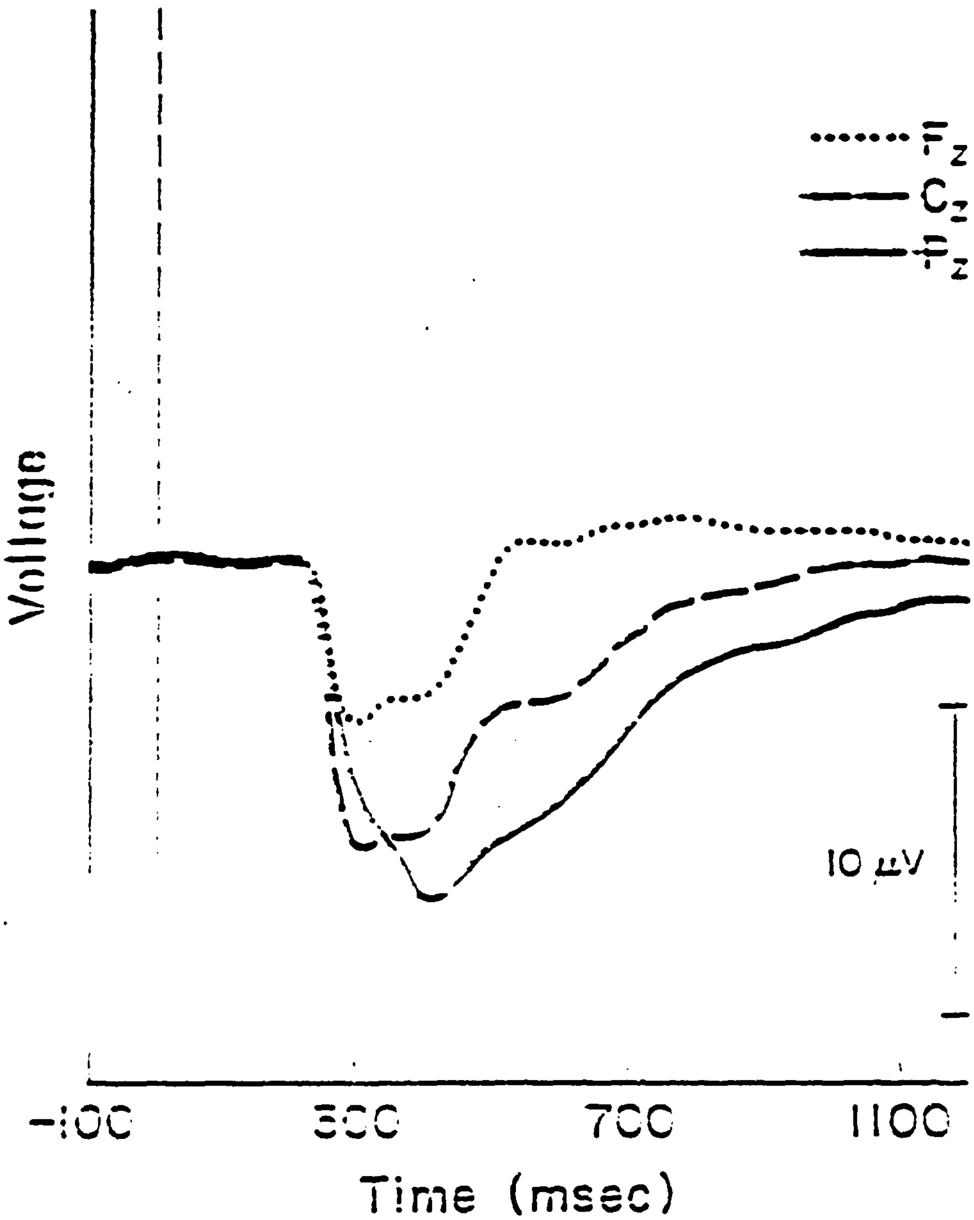
Figure 5
Grand Average ERP for the Single Task Tone Oddball



These waveforms represent the scalp distribution of the ERP averaged across all subjects elicited by targets in the single task auditory oddball condition.

Figure 6

Average ERP for the Single Task Step Count Condition



These vaveforms represent the scalp distribution of the ERP averaged across all subjects elicited by targets in the single task visual oddball contions.

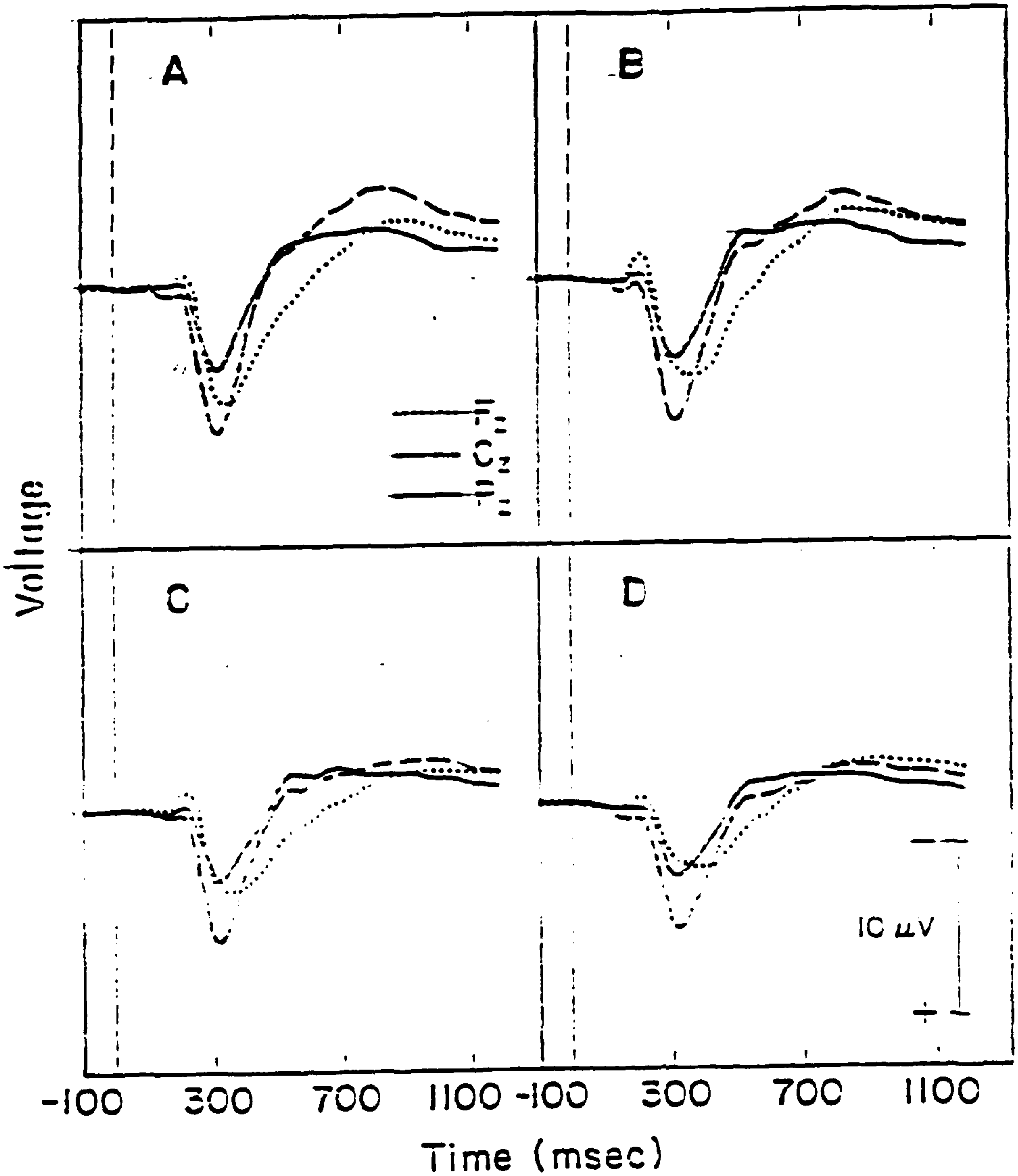
[F(1,39)=190.93, p<.01]; and displayed enhanced positivity for targets at Pz [F(2/78)=17.82,p<.01]. Because this component also occurred in the appropriate latency range, it is identified as the P300. Thus, the visual task employed in this experiment was also associated with typical P300s. However, visual inspection of the waveforms indicates that the P300 overlaps an earlier component maximal at Cz (producing a double peak at Pz). The overlapping component was also extracted by the PCA which confirmed its central maximal distribution and indicated that it was not larger for targets than non-targets. Although this component is clearly not a P300, the fact that it overlaps the P300 posed problems for the analysis of the primary task waveforms which will be elaborated below.

In summary, both the visual and auditory stimuli utilized in the subsequent step tracking conditions were associated with P300s occurring in the traditional latency range and with the expected scalp distribution under standard single task oddball conditions. For these reasons, it may be assumed that the waveforms associated with these stimuli in the various single and dual tracking conditions also contain traditional P300s.

Primary Task ERP Data Since the tone and step stimuli utilized for the primary and secondary tasks did elicit well defined P300s, the single and dual tracking task ERP data will now be examined to determine how the pattern of decrements evident in the overt performance data is related to variations in the amplitude of the P300 components elicited during the various single and dual tracking tasks.

The waveforms associated with step changes during the various dual task tracking conditions are displayed in Fig. 7. It is evident that compared to the single task visual oddball, in which step changes were counted, primary task P300 amplitudes in all the conditions were greatly reduced. Although

Figure 7
Grand Average ERPs Elicited by Dual-Task Step Changes



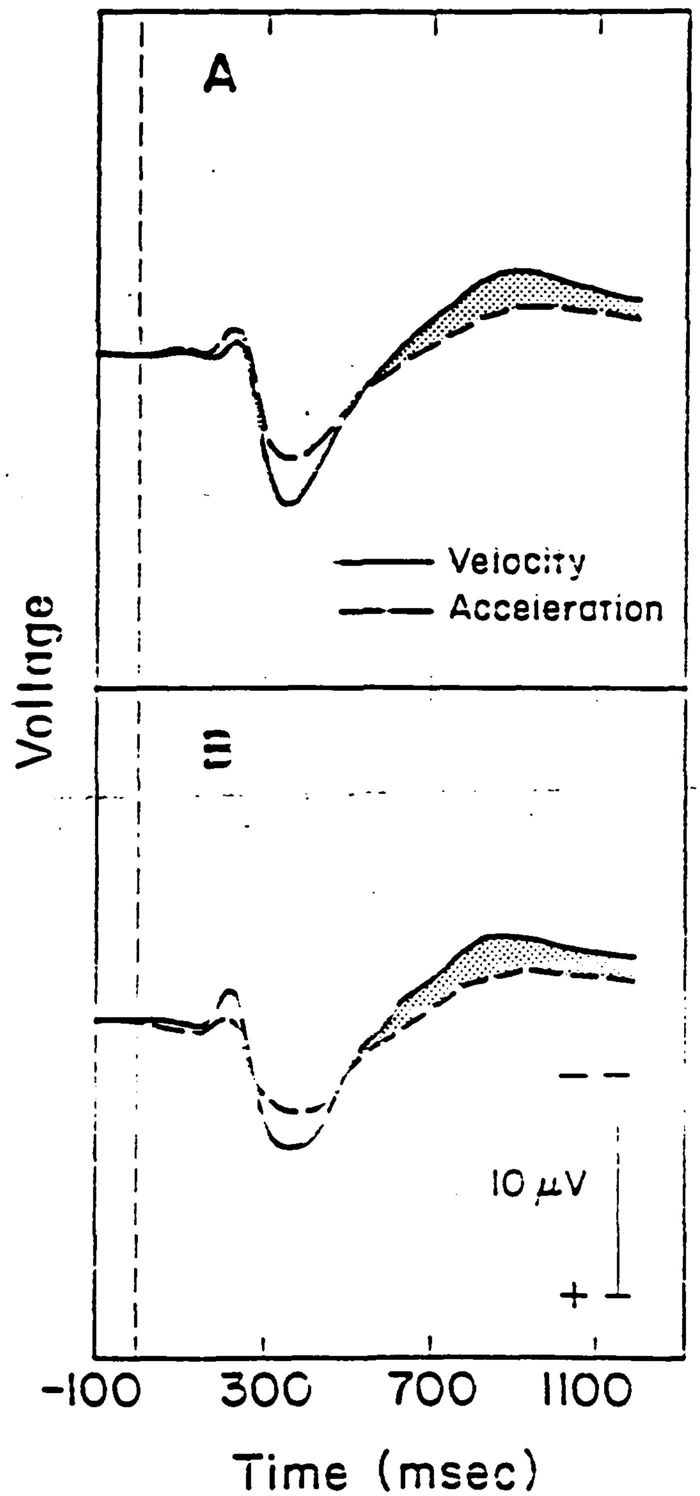
The scalp distribution of the ERPs elicited by the primary task step changes averaged across all subjects, is illustrated for the various step tracking conditions. A Velocity tracking in one dimension. B Velocity tracking in two dimensions. C Acceleration tracking in one dimension. D Acceleration tracking in two dimensions.

differences between conditions can be seen with regard to the amplitude of the large central maximal positive deflection which dominates the waveforms even at Pz, this peak is too early in latency and has the wrong scalp distribution to be identified as the P300. The P300 in these waveforms overlaps with this component producing differential returns to baseline for the different conditions.

Fig. 8 displays the effect of system order upon the primary task waveforms in both one and two dimensions. The cross-hatched areas indicate regions of increased positivity due to increased system order. However, the differences evident in the superaverages are small, presumably due to overlap with the earlier Cz maximal component. Because of this component overlap, a more detailed discussion of P300 amplitude will be given after the presentation of the PCA results.

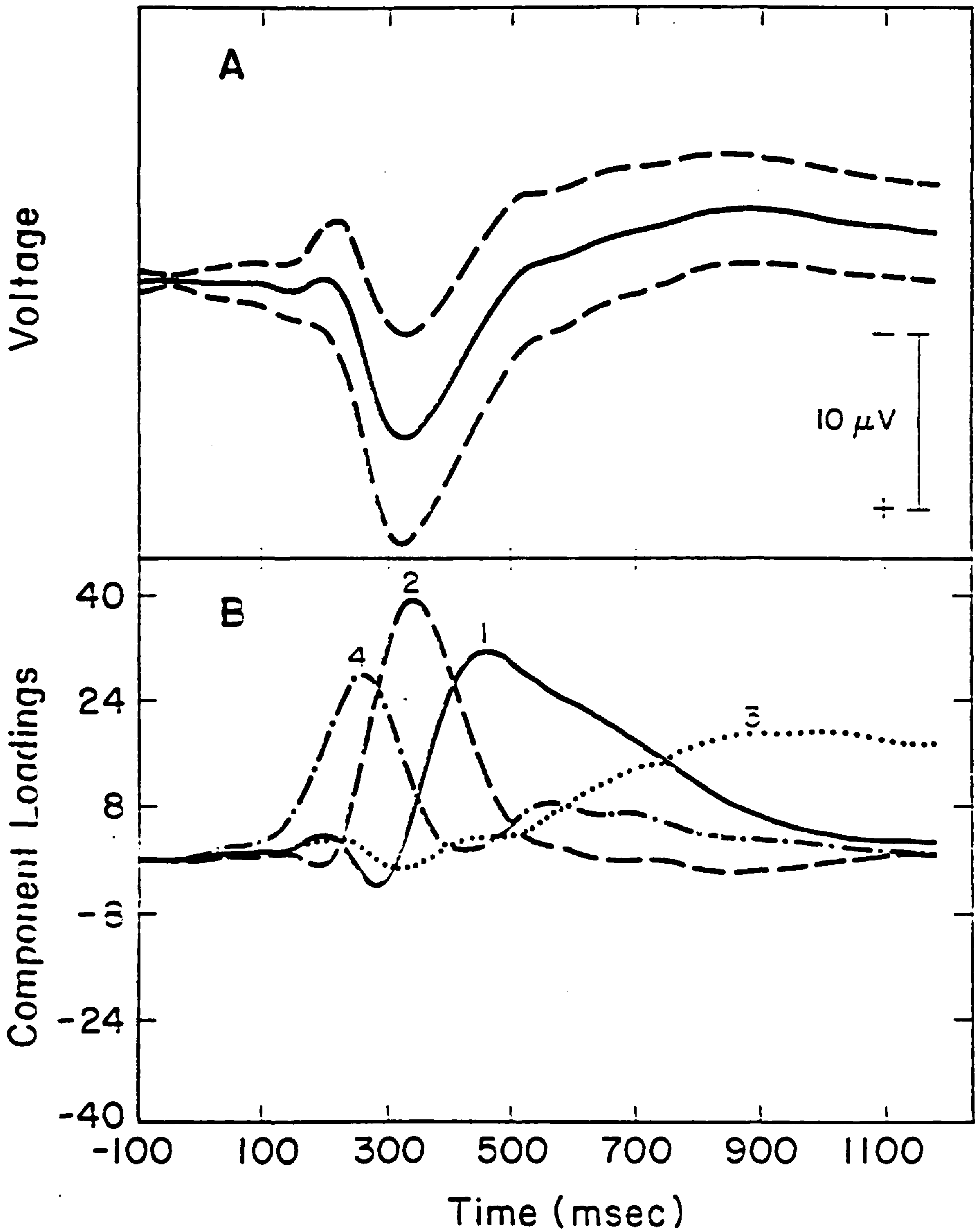
To facilitate comparisons between single and dual primary task ERPs, a single PCA was performed on the waveforms associated with the single and dual-task conditions. Such a comparison should validate the evidence provided by the RMS error data which indicates that subjects did indeed protect there primary task performance. Thus, if subjects did in fact consider the tracking task to be primary, dual task P300 amplitudes should be comparable to single task P300 amplitudes in all of the conditions. In other words, the resources allocated to the primary task under dual task conditions should be comparable to the resources allocated during single task performance. The data matrix submitted to the PCA consisted of 960 trials (40 Subjects x 2 Task levels x 2 Dimensions x 2 Control Orders x 3 electrodes), and four of the components extracted were Varimax rotated. The component structure extracted by this PCA is displayed in Fig. 9.

Figure 8
The Effect of Increased System Order on Parietal ERPs



The effect of increased system order upon the parietal waveforms is shown. The cross-hatched areas indicate increased positivity as a function of increased system order. A. One dimensional tracking. B. Two dimensional tracking.

Figure 9
Grand Mean Waveform and Component Structure of Primary Task ERPs



The output of the PCA performed on the data from the single and dual tracking tasks is displayed. A) The grand mean waveform (+/- 1 S.D.). B) The loadings of the four components extracted on each time point.

Component 1 is active in the appropriate latency range, and with the correct parietal maximal scalp distribution [$F(1,39)=219.07, p<.01$] to enable its identification as the component corresponding to P300. Overall, primary task P300 amplitude increased both as a function of increasing the number of dimensions [$F(1,39)=6.20, P<.05$] as well increasing the control order [$F(1,39)=33.32, p<.01$] of the tracking task with no significant interaction. Furthermore, both the dimension and order effects interacted with electrode site such that modulation of the component was greater at Cz [$F(2,78)=7.13, p<.01$; and $F(2,78)=28.13, p<.01$, respectively] even though the component loaded maximally on the Pz electrode as noted above.

Although the requirement to track in dual as opposed to single task conditions did not significantly affect P300 amplitude (as evidenced by the lack of a main effect of task level), the interactions of Task level x Dimension x Electrode [$F(2,78)=5.61, p<.01$] and Task level x Order x Electrode [$F(2,78)=11.09, p<.01$) were both significant, indicating that some differences between the single and dual task conditions existed. The presence of these task level differences requires that the effects of system order and dimensionality be analyzed separately for the single and dual tracking conditions.

Because the results of a PCA can be influenced by latency jitter between conditions, a base to peak analysis of primary task P300 amplitude was attempted at this point. However, base-to-peak estimates of primary task P300 amplitude could not be obtained for several reasons. As can be seen by comparing the waveforms, the visible P300 peak present in the single task step count oddball condition is greatly reduced in the single and dual-task tracking conditions. Thus, there was no clear inflection point at any electrode site. The difficulty in picking a peak is assumed to result from

the fact that not only was P300 amplitude greatly reduced as a function of introducing the tracking task; but also because of the overlap with the earlier Cz maximal component number 2. This component displayed the opposite sensitivity to experimental manipulations. However, the fact that the P300 was sensitive to the experimental manipulations in the predicted fashion, even though the overlapping component responded in the opposite manner, is evidence that robust P300s were present in the waveforms.

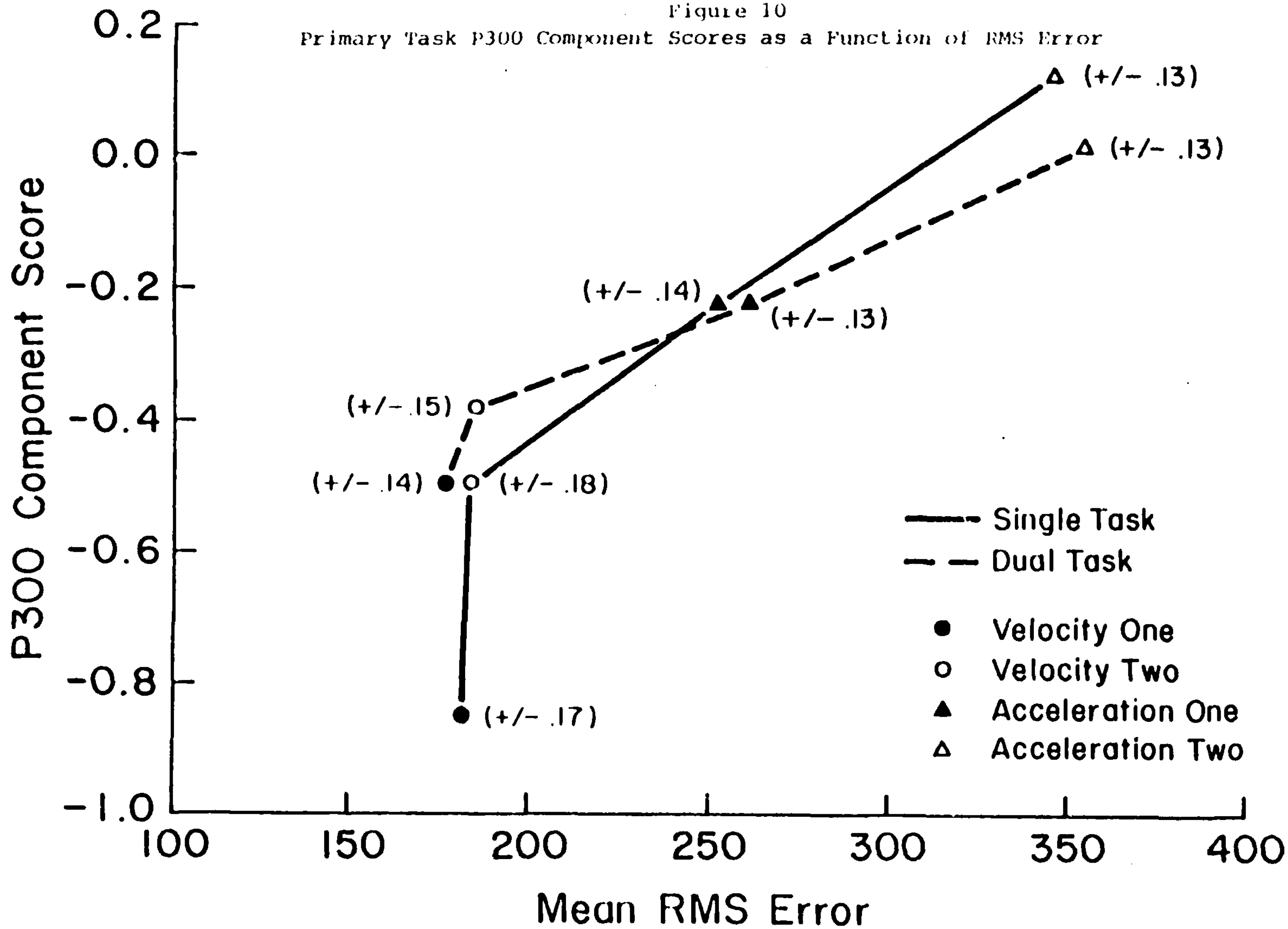
Because of the inability to obtain base-to-peak primary task measures, and because in both single and dual-task conditions the amplitude differences were significantly greater at Cz, primary task P300 amplitude differences were assessed by comparing the average component scores at Cz obtained by the PCA outlined above for the various tracking conditions. The relevent mean component scores are presented in Fig. 10. In most cases, the dual task P300 amplitudes are greater than or equal to the single task P300 amplitudes indicating that the subjects did indeed adopt satisfactory policies of resource allocation under the dual task conditions. Only in the most difficult condition was the dual task P300 amplitude slightly smaller, and this difference was not, in fact, significant ($p>.10$).

The main difference between the single and dual task amplitudes is in the easiest one dimensional velocity tracking conditions. Here subjects consistently produced very low amplitude P300s during single task performance and somewhat greater P300 amplitudes during the corresponding dual task condition. The data clearly demonstrate, however, that as the difficulty of the primary task was increased, the amplitude of the primary task P300s also increased.

Secondary task ERP data Since the tones utilized by the secondary task were clearly capable of eliciting P300s (as indicated by the large P300 present

Figure 10

Primary Task P300 Component Scores as a Function of RMS Error



The scores for the P300 component are displayed for the various step tracking conditions. Higher component scores reflect increased P300 amplitudes.

in the single task oddball data presented above), the secondary task data will now be examined to determine the extent to which variations in primary task workload (manifested in the RMS error, subjective effort ratings, and primary task ERP data) are reflected in amplitude variability associated with the various secondary task oddballs. In addition, the single and dual task oddball data will be compared to see the overall effect upon P300 amplitude due to the imposition of the tracking task.

The grand average ERPs for targets which were associated with the various dual task tone count conditions are displayed in Fig. 11. The large P300s present in the single task count conditions are greatly reduced in amplitude during the dual task conditions. This indicates that even during the simplest tracking conditions, the tracking task consumed a large proportion of the resources allocated to the counting task in isolation. However, though the P300 is small in all of the dual task conditions, the amplitudes are not equal in all tasks. As predicted, the one-dimensional velocity condition was associated with the largest secondary task P300, and the smallest secondary task P300 was produced by the most difficult two-dimensional acceleration condition (see Fig. 12).

Because visual inspection of the waveforms for individual subjects indicated some latency variability, and to facilitate amplitude comparisons between the single and dual-task auditory oddballs, pairwise comparisons of P300 amplitude differences in the various conditions were performed on base-to-peak estimates of P300 amplitude rather than on amplitude estimates based on the component loadings from a PCA. Had it been possible to obtain base-to-peak estimates for the primary task ERPs the same procedure would have been followed for the analysis of primary task P300s.

between the allocation of processing resources to the two tasks was presumed to exist, a reciprocal relationship between primary and secondary task P300 amplitudes was predicted. This prediction of primary and secondary tas P300 amplitude reciprocity was confirmed in all of the conditions in which it was tested. Thus, this experiment further illustrates the utility of the P300 as a tool to aid in the decomposition of mental workload.

In addition to validating the prediction of P300 amplitude reciprocity, this experiment produced a number of ancillary findings that we will now proceed to enumerate. In particular, previous findings concerning the nature of the manipulations of system order and dimensionality obtained in this laboratory have been both replicated and extended. Thus, the conclusion by Wickens et. al. (1983) that the manipulation of system order during a one-dimensional tracking task produces a salient drain on central/perceptual processing resources has been confirmed and extended to the two dimensional case. Secondary task P300s declined and primary task P300s increased in amplitude as a function of increased system order in both one and two dimension.

The finding of Wickens et. al. (1977) indicating the lack of a dimensionality effect upon P300 amplitude in velocity systems was also confirmed. Primary and secondary task P300 amplitude did not significantly change as a function of this manipulation when subjects were tracking with a velocity control system. However, the demand for central/perceptual resources did increase as a function of the dimensionality manipulation when subjects were tracking with an acceleration control system.

In conclusion, this experiment provides the first evidence from concurrently recorded primary and secondary tasks, that there is a reciprocal relationship between the amplitudes of the P300s associated with the two tasks. Furthermore, this relationship was investigated and confirmed under a variety of levels of primary and secondary task competition for processing resources. The results were intrepretated within a multiple resources model of dual task performance in which the allocation of processing resources to the two tasks was presumed to determine primary and secondary task P300 amplitude. Thus, because a reciprocal relationship

It would also be interesting to examine P300 amplitude reciprocity in conditions where primary and secondary task stimuli could occur simultaneously. It will be recalled that in this experiment, presentations of the oddball stimuli were constrained so that the recording epochs would not overlap with those of the step changes. At least two different outcomes can be predicted. Either the subjects will allocate all of their attention to the primary task, in which case the resulting ERP should resemble the primary task ERP under single task conditions; or they will attempt to divide their attention between the two tasks, in which case the resulting ERP may display a mixture of primary and secondary task processing.

A further issue concerns the extent to which the competition for perceptual resources can be localized to earlier stages of processing. A slight modification of the oddball task may provide insight into this question. A large body of research has indicated that the early negative components of the ERP are sensitive to channel selections in selective attention tasks (Hansen and Hillyard, 1980; Harter and Previc, 1978; Hillyard, Hink, Schwent, and Picton, 1973; Hink and Hillyard, 1976; Nataanen, 1975; Nataanen and Michie, 1979). The secondary task employed in this experiment could be easily modified to embody the selective attention paradigm. For example, tones could be presented monaurally to the left and right ears and subjects could be instructed to count only high pitched tones presented to one ear. Hillyard's data indicate that both the targets and the non-targets presented on the attended channel display enhanced negativity (termed the Nd component) at a latency within the first 100 msecs. of the epoch. It would be interesting to determine whether the channel selections indexed by the Nd can be affected by increasing the difficulty of a concurrently performed primary tracking task.

processing, a visual oddball task could have been employed. Although some preliminary work investigating the effects of overlapping primary and secondary task stimulus integrality has been carried out (Kramer, Wickens, and Donchin, in preparation), further research is needed to determine the extent to which P300 amplitude decrements associated with secondary tasks of different modalities can be compared to assess workload.

For all these reasons, P300 amplitude measured under dual-task conditions appears to be an important tool to aid in the analysis of demands placed upon operators in man-machine systems. In terms of Wickens' (1981) five criteria for the ideal workload metric, the P300 appears to be an unobtrusive measure sensitive to graded changes in task difficulty. Furthermore, the P300 appears to be diagnostic of perceptual as opposed to response-related processing. Finally, it is conceivable that with further refinements, such as the application of step-wise discriminant analysis techniques (Donchin and Herning, 1975), the bandwidth and reliability of the P300 may be of sufficient quality to permit the analysis of workload on a moment by moment basis; however, further research is needed to assess the feasibility of this idea.

In fact, a number of questions for future research are suggested by this study. The particular step tracking task employed here as the primary task did not, unfortunately, produce large P300s. This is not surprising given that all the step changes constituted equiprobable targets. It would be desirable, therefore, to investigate P300 amplitude reciprocity under conditions which would elicit larger primary task P300s. Presumably a manipulation of probability, possibly combined with the introduction of a target vs. non-target step change dimension, would produce this effect.

execution processing, it can be argued that the workload demands indexed by P300 amplitude can be localized to central/perceptual stages of processing. Thus, Wickens' claim (Wickens et. al., 1980) that the manipulation of system order requires not only increased response selection and execution processing but also increased central/perceptual processing is validated by the fact that the acceleration tracking conditions were associated with increased primary task and decreased secondary task P300 amplitudes when compared with the velocity tracking conditions.

The validation of P300 amplitude as a metric of a particular aspect of the workload demands of a task has a number of theoretical and applied implications. As mentioned earlier, the oddball task is an attractive secondary task for a number of reasons. The most important of these reasons is that an oddball task can be applied in a relatively non-obtrusive fashion in many different situations because there is no need for an overt response. Thus, because subjects can count the oddball stimuli rather than respond overtly to them, competition for response related processing resources is reduced. The fact that performance of the oddball task can be measured by a psychophysiological measure such as the P300 is the single greatest advantage of the oddball paradigm. The RMS error data from this experiment confirm that, indeed, a secondary oddball task can be imposed as a dual task with only a minimal cost to the performance of the primary task.

Another advantage of the oddball task is that oddball stimui of different modalities can be used to elicit P300s. The modality of the secondary task can, therefore, be chosen to eliminate competition for modality specific processing resources. In this experiment, an auditory secondary task was chosen because the step-tracking task required visual stimulus processing. Had the primary task relied more upon auditory

secondary task P300 amplitudes should decline as a result of the drain upon this limited commodity.

The data collected during this experiment confirm this assertion. As the difficulty of the primary task increased, the RMS error measures also increased. Furthermore, the amplitude of the P300s associated with primary task step changes increased, while the amplitude of the secondary task P300s elicited by the auditory stimuli decreased in the predicted fashion. Additionally, the theory of resource reciprocity derives futher suppport from the fact that large increases in primary task P300 amplitudes as a function of task difficulty were accompanied by large decreases in secondary task P300 amplitudes; while small increments in primary task P300 amplitudes were associated with small decrements in secondary task P300 amplitudes.

The fact that primary task P300s increased in amplitude only in conditions where secondary task P300s decreased provides convincing evidence that P300 amplitude reflects the allocation of processing resources. Thus, had primary task P300s increased as a function of increased dimensionality in velocity systems while secondary task P300s remained the same, the construct of reciprocity would have been in jeopardy. However, this was not the case. In all conditions, the increase in primary task P300 amplitude was proportional to the decrease in secondary task P300 amplitude. An examination of figure 14 confirms that the summation of primary and secondary task P300 amplitudes yields a constant value.

The utility of P300 amplitude as a metric of mental workload is, therefore, confirmed by this study. Furthermore, the P300 is a metric of a particular type of mental workload. Because the subset of processes upon which the P300 is dependent are sensitive to manipulations of stimulus evaluation but are relatively independent of response selection and

Conclusions

This experiment provides strong empirical support for the predicition of a reciprocal relationship between the amplitudes of the P300s associated with two concurrently performed tasks. This predicition is derived from a large body of evidence (reviewed above) which has indicated that variations in P300 amplitude are sensitive to the manner in which subjects allocate processing resources between two tasks under dual-task conditions. In other words, P300 amplitude has emerged as a psychophysiological metric of mental workload.

Thus, when subjects are instructed to protect performance of one task (the primary task) at the expense of another task (the secondary task) it is assumed that a greater allocation of resources to the primary task is required to maintain performance when the difficulty of this task is increased. If the P300 is indeed sensitive to the allocation of processing resources, the pattern of primary and secondary task P300 amplitude variability should reflect this reallocation of resources to the primary task.

The RMS error data (as well as the subjective ratings of effort) confirm that the orthogonal manipulation of system order and dimensionality employed by this study successfully produced a wide variability in performance within which to assess the reciprocity of primary and secondary task P300 amplitudes. Because the difficulty of the secondary task was held constant during all the step-tracking conditions, the model of resource reciprocity upon which this experiment is based predicts that as the tracking task is made more difficult, primary task P300 amplitudes should become larger, due to the allocation of additional processing resources; and

Table 3

Mean Slope and Intercept Values for Individual Reciprocity Functions

Sub.	Slope	Intercept	Sub.	Slope	Intercept
1	0.05	96.62	21	1.53	54.72
2	-0.11	127.18	22	0.40	85.66
3	0.70	84.44	23	-1.12	134.72
4	0.33	87.97	24	-0.98	116.76
5	0.28	51.88	25	0.05	89.43
6	0.05	97.55	26	-0.58	151.54
7	0.23	70.44	27	0.55	56.62
8	0.30	87.83	28	-0.20	130.31
9	0.43	83.36	29	-0.35	90.99
10	-1.23	151.67	30	0.17	65.31
11	-0.06	93.17	31	-0.25	107.22
12	1.40	43.04	32	-0.86	138.12
13	0.00	129.88	33	0.93	71.70
14	-0.13	101.28	34	0.70	57.01
15	-1.31	132.43	35	0.23	98.31
16	-0.63	137.22	36	-0.45	121.85
17	-0.20	85.12	37	1.14	65.18
18	0.21	80.94	38	0.65	69.87
19	0.36	73.45	39	0.04	93.99
20	-0.11	117.01	40	0.34	72.72

Mean slope = 0.06 Std. error = 0.10

Mean intercept = 95.48 Std. error = 5.44

Because the amplitude measures were derived from different metrics, the measures were submitted to the same range correction transformation outlined above. The line in the center of the graph represents the sum of the single and dual task curves. Perfect amplitude reciprocity would generate a function with a slope of zero and an intercept value of 100. As can readily be seen by examining the actual average, the evidence for amplitude reciprocity is quite good. Difficult tracking conditions produced a demand for perceptual resources resulting in increased primary task P300s and decreased secondary task P300s. Furthermore, the greater the increase in primary task P300 the greater the decrease in secondary task P300. This experiment, therefore, provides the first evidence for amplitude reciprocity obtained from concurrently recorded primary and secondary tasks of different modalities.

To determine the extent to which this pattern of reciprocity held true on a single subject basis, separate reciprocity functions were obtained for each subject and the regression line for these functions were computed. If the single subjects also demonstrated signifcant reciprocity the mean slope of these derived functions should equal zero and the mean intercept should equal 100. These data are presented in Table 3. Although there was significant variability within the subjects (indicating the presence of instances of both under and over reciprocity) the obtained value of 0.06 for the mean slope did not differ significantly from the predicted value of 0 ($t=0.10, p>0.10$); and the mean intercept value of 95.48 did not differ significantly from the predicted value of 100 ($t=0.13$, $p>0.10$). Thus, evidence in support of the reciprocity theory was obtained both across and within subjects.

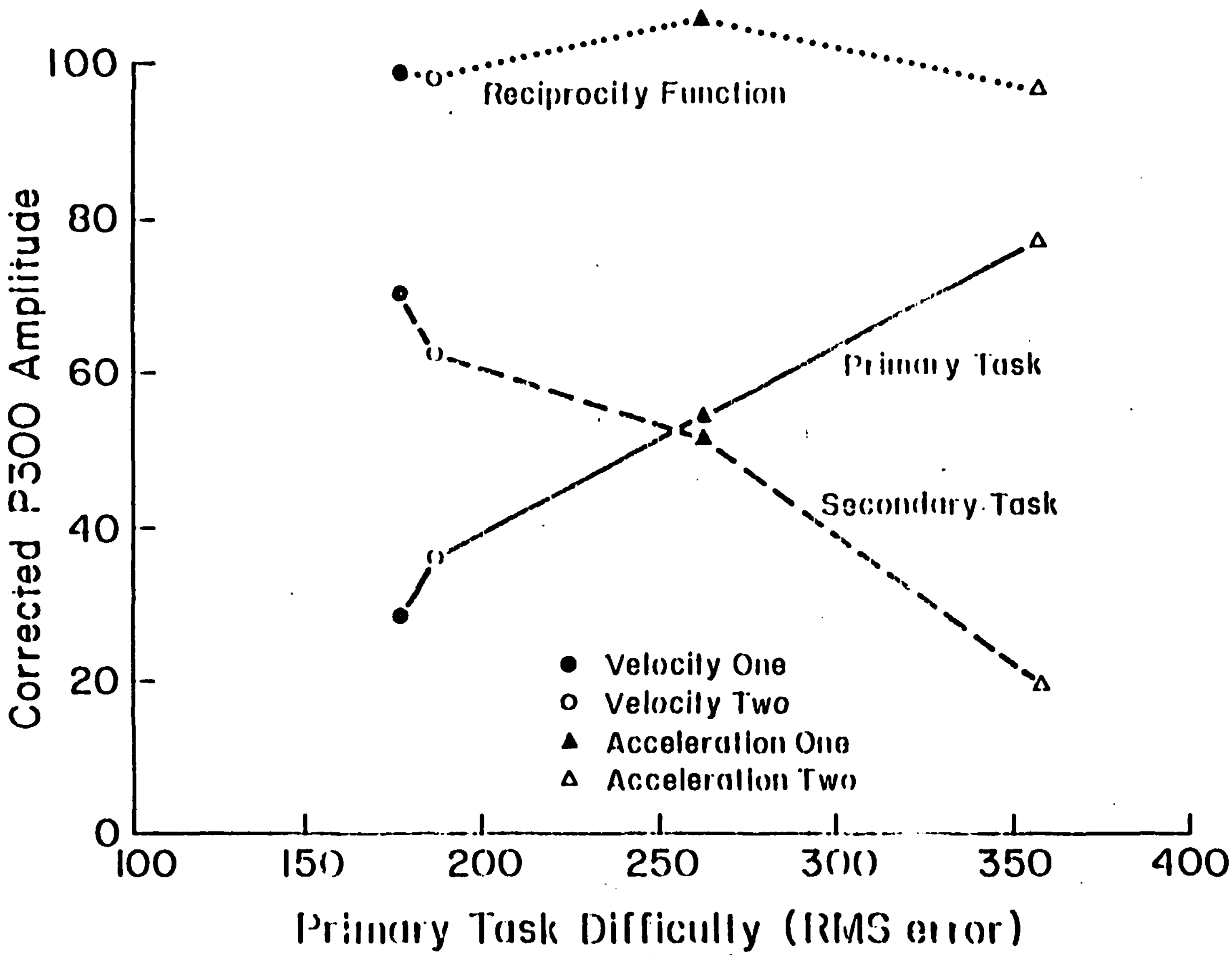


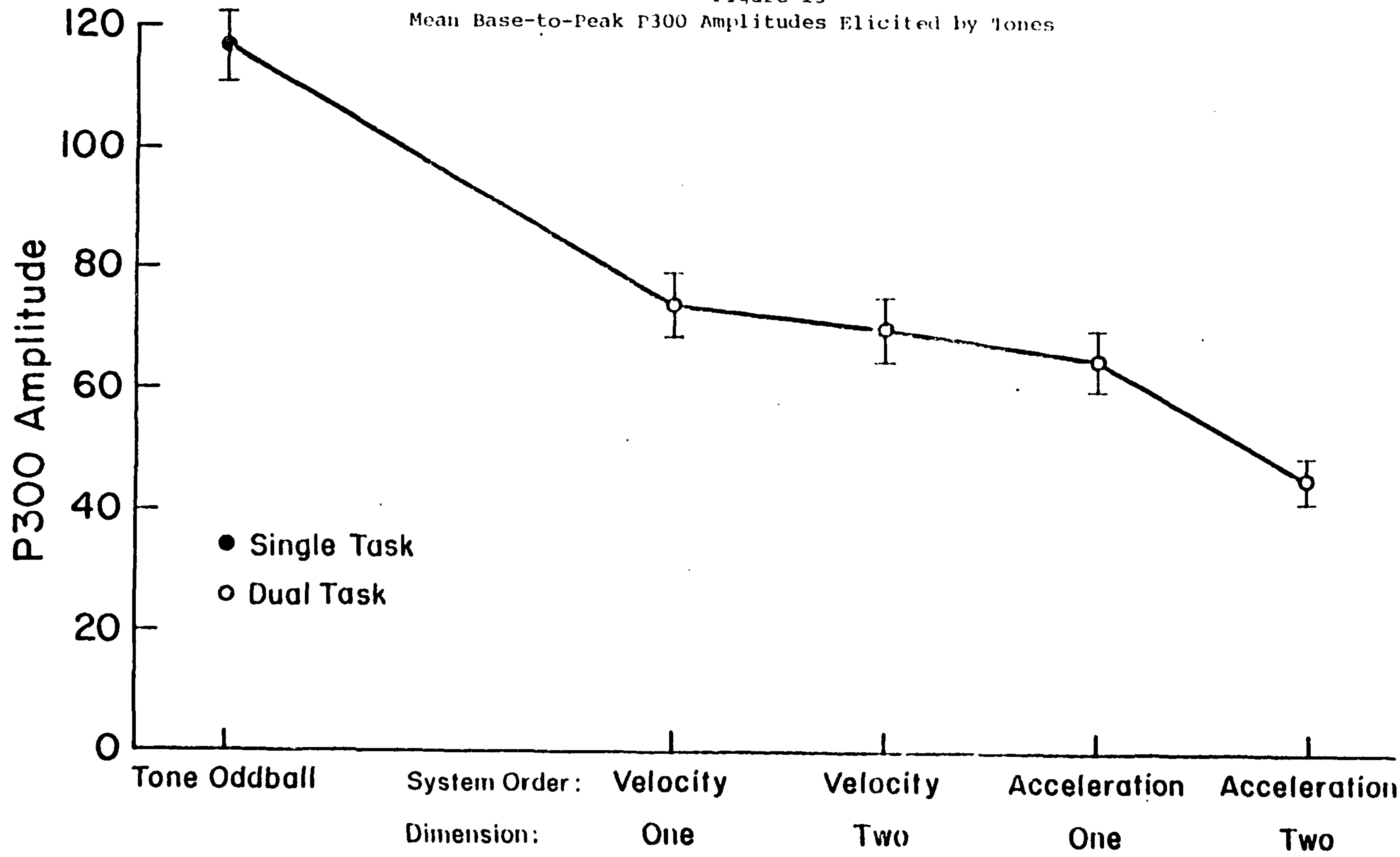
Figure 14

The Reciprocity of P300 Amplitude

The range corrected primary and secondary task P300 amplitudes for all of the dual task tracking conditions are plotted as a function of RMS error score. The reciprocity function represents the sum of the primary and secondary task amplitude measures.

Figure 13

Mean Base-to-Peak P300 Amplitudes Elicited by Tones



Mean base-to peak estimates of P300 amplitude averaged across all subjects for each single and dual task auditory oddball is shown in arbirary units (one hundred units = approx. ten microvolts).

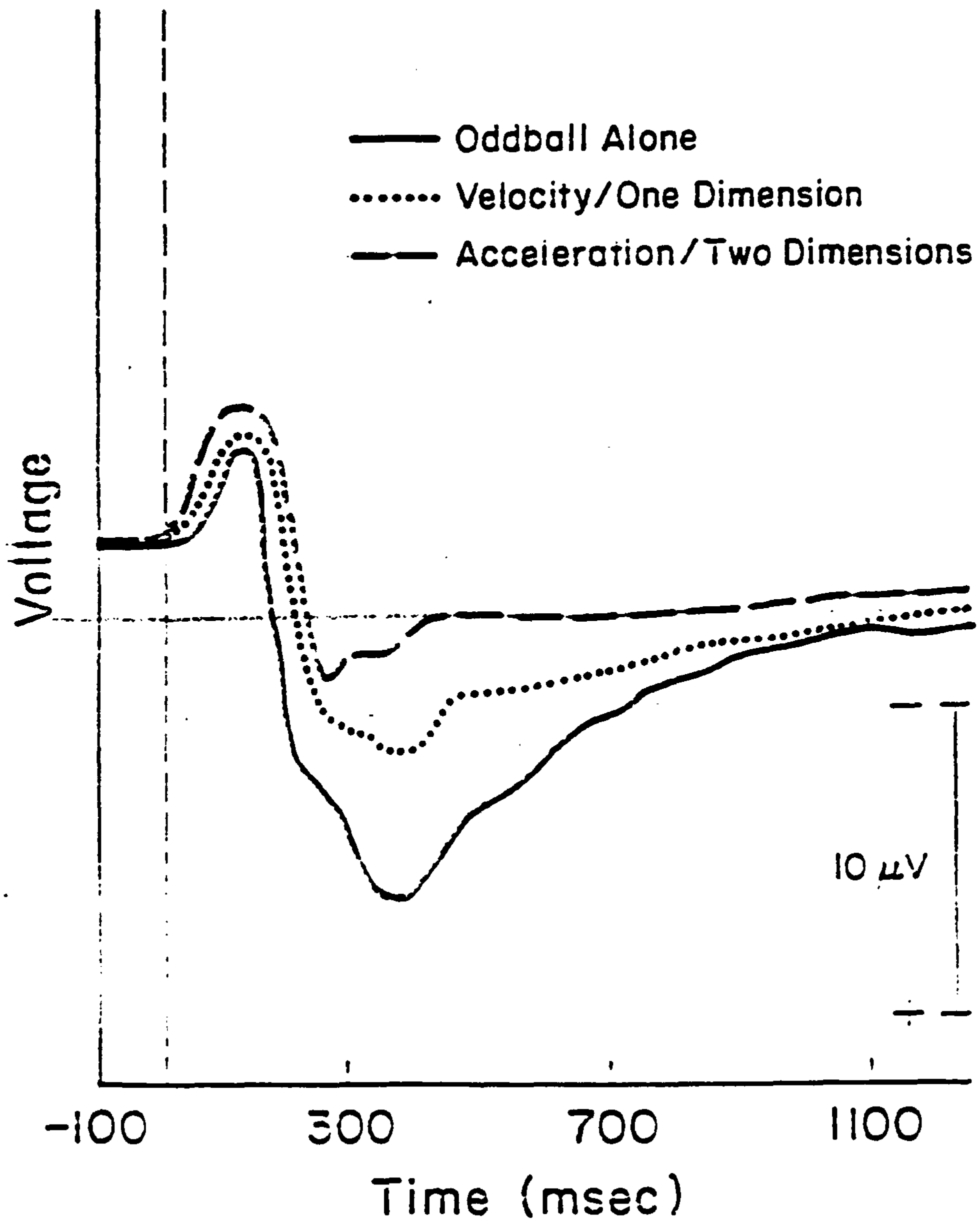
Thus, for the analysis of secondary task P300 amplitude, a time window which included the P300 peak was chosen for each subject, and a peak picking algorithm was developed to select the maximum positive inflection point at the parietal electrode within this window. The average digitized voltage of the pre-stimulus 100 msec. baseline at Pz was then subtracted from the digitized value associated with the P300 peak at Pz.

Base-to-peak amplitude measures were obtained for each single and dual-task oddball condition. Mean base-to-peak amplitudes from the various tone count conditions are presented in Fig. 13. ANOVAs performed on the dual-task base-to-peak measures confirm that P300 amplitude varied as a function of both the dimensionality [$F(1,39)=14.99, p<.01$] and system order manipulations [$F(1,39)=21.01, p<.01$]. However, a significant dimension x order interaction [$F(1,39)=7.79, p=.01$) was also evident.

Because of this interaction, the significance of subsequent pairwise comparisons was evaluated by means of Tukey tests. Acceleration systems were associated with decreased P300 amplitudes regardless of whether subjects were tracking in one or two dimensions (p levels $< .01$). However, tracking in two dimensions rather than one was associated with secondary task P300 amplitude decrements only when subjects were tracking in acceleration systems ($p<.01$). This pattern of results is in complete harmony with the predicted outcome. Thus, as primary task difficulty was increased, larger P300 amplitudes were associated with the primary task and smaller P300 amplitudes were associated with the secondary task.

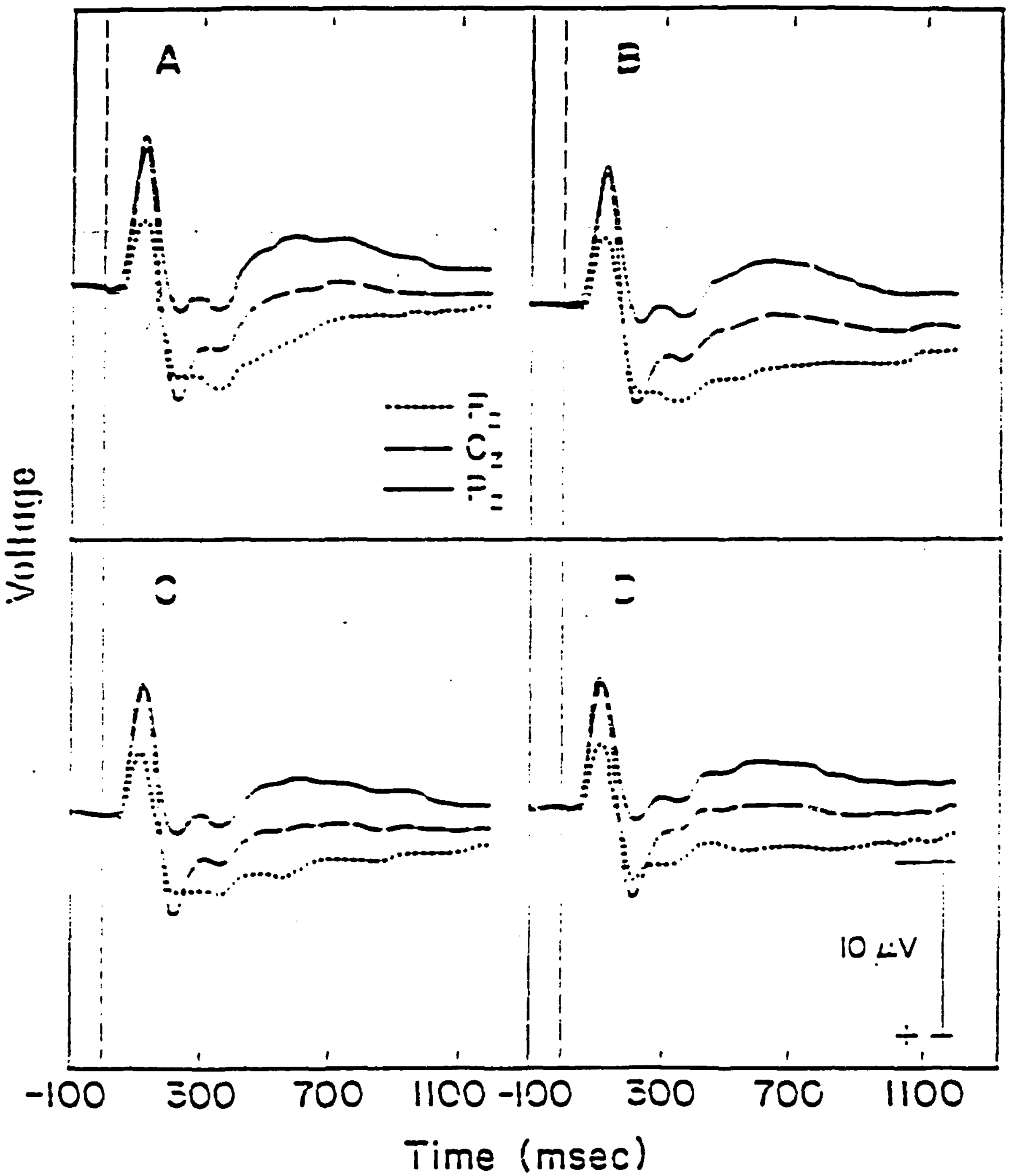
Reciprocity confirmation The dual task psychophysiological data provide important evidence concerning the nature of P300 amplitude reciprocity and its relationship to mental workload. Fig. 14 shows both the primary and secondary task P300s plotted on the same scale as a function of RMS.

Figure 12
Parietal ERPs for Selected Single and Dual Task Tone Oddballs



The parietal ERP associated with the single task auditory oddball as well as the secondary task waveforms associated with the easiest (velocity in one dimension) and the most difficult (acceleration in two dimensions) tracking conditions are shown.

Figure 11
Grand Average ERPs elicited by Dual Task Tones



These waveforms represent the scalp distribution of the ERP averaged across all subjects elicited by secondary task target tones during the various dual task tracking conditions. A) Velocity tracking in one dimension. B) Velocity tracking in two dimensions. C) Acceleration tracking in one dimension. D) Acceleration tracking in two dimensions.

References

Allport, D.A., Antonis, B., and Reynolds, P. On the division of attention: A disproof of the single channel hypothesis. Quarterly Journal of Experimental Psychology., 1972, 24, 255-265

Berlyne, D.E. Conflict, Arousal, and Curiosity. New York: McGraw-Hill, 1960.

Borg, G. Subjective aspects of physical and mental load. Ergonomics, 1978, 21, 215-220.

Brown, I.D. Dual task methods of assessing workload. Ergonomics, 1978, 21, 221-224.

Derrick, W.L. The relationship between processing resources and subjective dimensions of operator workload. In R. Sugarman (Ed.). Proceedings 25th Annual Meeting of the Human Factor Society., Santa Monica: Human Factors, 1981.

Donchin, E., & Cohen, L. Averaged evoked potentials and intra-modality selective attention. Electroencephalography and Clinical Neurophsiology. 1967, 22, 537-546.

Donchin, E. & Heffley, E. Minicomputers in the signal averaging laboratory. American Psychologist., 1975, 30, 299-312.

Donchin, E. & Herning, R.I. A simulations study of the efficacy of stepwise discriminant analysis in the detection and comparison of event-related potentials. Electroencephalography and Clinical Neurophysiology, 1975, 38, 51-68.

Donchin, E. & Gopher, D. What is workload -- An introduction, in L. Kaufman and K. Boff (Eds.), Handbook of perception and human performance. New York: John Wiley (in press).

Donchin, E. & Isreal, J.B. Event-related potentials -- Approaches to cognitive psychology. In R.E. Snow, P.A. Federico, & W.E. Montague (Eds.) Aptitude, learning, and instruction: Cognitive process analysis. Hillsdale, N.J.: Erlbaum, 1979.

Donchin, E., Kramer, A., & Wickens, C.D. Probing the cognitive infrastructure with event-related brain potentials. Proceedings of the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics. Edwards Air Force Base. California, 1982.

Donchin, E., Kubovy, M., Kutas, M., Johnson, R., Jr., & Herning, R.I. Graded changes in evoked reponse (P300) amplitude as a function of cognitive activity. Perception & Psychophysics, 1973, 14, 319-324.

Donchin, E., Ritter, W., & McCallum C. Cognitive psychophysiology; The endogenous components of the ERP. In E. Callaway, P. Tueting, $ S. Koslow

(Eds.), Brain event-related potentials in man. New York: Academic Press, 1978, pp. 349-441.

Duncan-Johnson, C.C. & Donchin, E. On quantifying surprise: The variation in event-related potentials with subjective probability. Psycholphysiology, 1977, 14, 456-467.

Duncan-Johnson, C.C. & Donchin, E. Series-based vs. trial-based determinants of expectancy and P300 amplitude. Psychophysiology, 1978, 15, 262.

Ellis, G.A. Subjective assessment pilot opinion measures. In A.H. Roscoe (Ed.) Assessing Pilot Workload. AGARD-A6-233, 1978.

Ford, J.M., Roth, W.T., & Kopell, B.S. Auditory evoked potentials to unpredicatable shifts in pitch. Psychophysiology, 1976, 13, 32-39.

Gratton, G., Coles, M. & Donchin, E. A new method for off-line removal of ocular artifact. Electroencephalography and Clinical Neurophysiology, 1983, 55, 468-484.

Hansen, J.C. & Hillyard, S.A. Endogenous brain potentials associated with selective auditory attention. Electroencephalography and Clinical Neurophysiology, 1980, 49, 277-290.

Harter, M.R. & Previc, F.H. Size specific information channels and selective attention: Visual evoked potential and behavioral measures. Electroencephalography and Clinical Neurophysiology, 1978, 45, 628-640.

Hillyard, S.A., Hink, R.F., Schwent, V.L. & Picton, T.W. Electrical signs of selective attention in the human brain. Science, 1973, 182, 177-180.

Hink, R.F. & Hillyard, S.A. Auditory evoked potentials during selective listening to dichotic speech messages. Perception and Psychophysics, 1976, 20, 236-242.

Howitt, J.S. Flight deck workload studies in civil transport aircraft. Nato AGARD Conference Proceedings, 1968, 56, 1-7.

Isreal, J.B., Chesney, G.L., Wickens, C.D., & Donchin, E. P300 and tracking difficulty: Evidence forl multiple resources in dual-task performance. Psychophysiology, 1980, 17, 259-273.

Isreal, J.B., Wickens, C.D., Chesney, G.L. & Donchin, E. The event related brain potential as an index of display monitering workload. Human Factors, 1980, 22, 211-224.

Jasper, H.H. The ten twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, 1958, 10, 371-375.

Johnson, R.E., Jr., & Donchin, E. On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography & Clinical Neurophysiology, 1978, 44, 424-437.

Johnson, R., Jr. & Donchin, E. Sequential expectancies and decision making in a changing environment: An electrophysiological approach. Psychophysiology, 1982, 19, 183-200.

Kahneman, D. Attention and Effort. Englewood, Cliffs, N.J.: Prentice Hall, 1973.

Knowles, W.B. Operator loading tasks. Human Factors, 1963, 5, 155-161.

Kramer, A.F., Wickens, C.D. & Donchin, E. An analysis of the processing requirements of a complex perceptual-motor task. Human Factors, 1983, 25, 597-622.

Kramer, A.F., Wickens, C.D. & Donchin, E. The processing of stimulus properties: Evidence for dual-task integrality. Submitted to Journal of Experimental Psychology: Human Perception and Psychophysics, in preparation.

Kutas, M., McCarthy, G., & Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197, 792-795.

Moray, N. Where is capacity limited? A survey and a model. In A.F. Sanders (Ed.), Attention and performance I, Amsterdam: North Holland, 1967.

Moray, N. Subjective mental workload. Human Factors, 1982, 24, 25-40.

Nataanen, R. Selective attention and evoked potentials in humans -- A critical review. Biological Psychology, 1975, 2, 237-307.

Nataanen, R. & Michie P.T. Early selective attention effects on the evoked potential. A critical review and reinterpretation. Biological Psychology, 1979, 8, 81-136.

Navon, D. & Gopher, D. On the economy of the human processing system. Psychological Review, 1979, 86, 214-255.

Norman, D., & Bobrow, D. On data-limited and resource-limited processes. Cognitive Psychology, 1975, 7, 44-64.

North, R.A. Task functional demands as factors in dual-task performance. Proceedings of the Twenty-first Annual Meeting of the Human Factors Society, San Fransisco: October, 1977.

Ogden, G.D., Levine, J.W., & Eisner, E.J. Measurement of workload by secondary tasks. Human Factors, 1979, 21, 529-548.

Rolfe, J.M. The secondary task as a measure of mental load. In W.T. Singleton, J.G. Fox, & D. Witfield (Eds.), Measurement of Man at Work. London: Taylor & Francis, 1971.

Roscoe, A.H. Assessing Pilot Workload. AGARD-A6-233, 1978.

Schneider, W. & Fisk, A.D. Concurrent automatic and controlled visual search: Can processing occur without cost? Journal of Experimental Psychology: Learning, Memory, & Cognition., 1982, 8, 261-278.

Shaffer, H. Multiple attention in continuous verbal tasks. In S. Dornick & P.M. Rabbitt (Eds.) Attention and Performance V. New York: Academic Press, 1975.

Sheridan, T.B. Mental workload: What is it? Why bother with it? Human Factors Society Bulletin, 1980, 23, 1-2.

Smith, A. Symbol Digits Modalities Test. Los Angeles: Western Psychological Services, 1973.

Sperling, G. & Melchner, M. Visual search, visual attention and the attention operating characteristic. In J. Requin (Ed.), Attention and Performance VIII. Hillsdale, N.J.: Erlbaum, 1978.

Sutton, S., Braren, M., Zubin, J. & John, E.R. Evoked-potential-correlates of stimulus uncertainty. Science, 1965, 150, 1187-1188.

Trumbo, D., Noble, M. & Swink, J. Secondary task interference in the performance of tracking tasks. Journal of Experimental Psychology, 1967, 73, 232-240.

Tukey, J.W. Exploratory Data Analysis. Reading, Mass.: Addison Wesley, 1977.

Vidulich, M. & Wickens, C.D. Time-sharing manual control and memory search: The joint effects of input and output modality competition, priorities, and control order. Engineering-Psychology Research Laboratory. University of Illinois at Urbana-Champaign. Technical Report EPL-81-4/ONR-81-4, December, 1981.

Wickens, C.D. The effects of divided attention in information processing and tracking. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 1-13.

Wickens, C.D. Workload: In defense of the secondary task. Personnel Training and Selection Bulletin, 1981, 2, 119-123.

Wickens, C.D. Processing resources in Attention. In R. Parasuramian, J. Beatty, & R. Davies (Eds.), Varieties of Attention. New York: Academic Press, 1983.

Wickens, C.D. Engineering Psychology and Human Performance. Columbus, Ohio: Charles Merrill Publishing Company, 1984.

Wickens, C.D., Derrick, W.D., Micalizi, J. & Berringer, D. The structure of processing resources. In G.E. Corrick, E.C. Haseltine &R.T. Durst (Eds.), Proceedings of the Human Factors Society, 24th Annual Meeting, Los Angeles, California, 1980.

Wickens, C.D., Isreal, J.B., Donchin, E. The event-related cortical potential as an index of task workload. In A.S. Neal & R.F. Palasek (Eds.), Proceedings of the Human Factors Society 21st Annual Meeting, Santa Monica, California, 1977.

Wickens, C., Kramer, A., Vanasse, L., and Donchin, E. Performance of concurrent tasks: A Psychological analysis of the reciprocity of information processing resources. Science, 1983, 221, 1080- 1082.

Wierwille, W.W. Physiological measures of aircrew mental workload. Human Factors, 1979, 21, 575-594.

Williges, R.C. & Wierwille, W.W. Behavioral measures of aircrew mental workload. Human Factors, 1979, 21, 549-574.

Yeh, Y., & Wickens, C. An investigation of the dissociation between subjective measures of mental workload and performance. Technical report EPL-84-1/NASA-84-1, 1984.

APPENDIX B13

P300 Latency and Reaction Time are Dissociated in a Sternberg Task

David L. Strayer, Demetrios Karis,

Michael G.H. Coles, & Emanuel Donchin

Manuscript in preparation, based on paper by David L. Strayer submitted in fulfillment of the first year requirements, Experimental Division, Department of Psychology, University of Illinois, 1984.

Appendix B13

P300 Latency and Reaction Time are Dissociated in a Sternberg Task

David L. Strayer

Cognitive Psychophysiology laboratory

University of Illinois

Running head: DISSOCIATION OF P300 LATENCY AND REACTION TIME

Abstract

Theios (1974) has reported sequential effects for repeated stimuli in the Sternberg task suggesting that the memory representations of the stimuli are rearranged on a trial by trial basis. If the P300 reflects the updating of working memory (Donchin, 1979,1981), then the P300 should reflect this restructuring. Twenty-five right handed males participated in the experiment. Subjects memorized 1,2,3,4, or 5 letters. 30 test stimuli were presented following each memory set. Subjects were to indicate whether or not the test stimulus was a member of the memory set. Subjects responded faster to positive stimuli than to negative stimuli. Furthermore, the larger the set size the slower the RT. The difference between the RTs for positive and negative stimuli remained constant across set size. The P300s elicited by positive stimuli occurred earlier than for negative stimuli. P300 latency increased as a function of set size for positive stimuli, however for negative stimuli P300 latency was relatively constant for set sizes 2-5. These data suggest that the RT response is a function of the same processes for both positive and negative stimuli, namely a serial comparison with the letters held in working memory. However, the P300 behaves differently for positive and negative stimuli. This suggests that the P300 is associated with a process which is elicited to maintain or establish the representation of the test stimulus in working memory for use on subsequent trials.

P300 Latency and Reaction Time are Dissociated in a Sternberg task

Sternberg (1966, 1967, 1969a, 1969b, 1975) developed a technique to infer the existence and properties of processing stages and the timing of mental events. The method assumes that reaction time (RT) is determined by the sum of the durations of a series of processing stages which are ordered such that one stage does not begin operating until the preceding stage has been completed. If the duration of a processing stage is affected by experimental manipulation, mean RT will increase at a constant rate with changes in the duration of this stage. If more than one processing stage is manipulated, mean RT will vary in an additive manner. That is, the total effect on RT will be equal to a linear summation of the effects that each manipulation exercises on RT. From this assumption it follows that when the effects of two experimental variables on RT are additive, they can be assumed to affect different processing stages.

Sternberg (1966) used this technique to study how information is retrieved from short-term memory. Subjects were required to memorize from one to six digits (hereafter referred to as the memory set) and then determine if a test stimulus was contained in the memory set. Mean RT increased linearly as a function of the number of elements presented in the memory set (subsequently referred to as set size). The linearity of the function was consistent with the hypothesis that the search through memory is executed as a serial comparison process. The test stimulus, according to this view, is compared sequentially to the representations of each of the items in the memory set. The comparison takes, on the average, a fixed duration per item. The slope of the regression line relating RT to the size

of the set was taken to represent the mean comparison time. The intercept of this regression line was interpreted to be the sum of the duration of other processes which are independent of set size.

Sternberg (1967, 1969a, 1969b) proposed a model in which RT is determined by, at least, four additive processing stages. First, the test stimulus is encoded. Second, a serial comparison process is invoked in which the test stimulus is successively compared to the elements in the memory set. Third, a binary (yes/no) decision is made. Finally, the appropriate response is selected and executed.

A surprising outcome of Sternberg's (1966) experiment derived from a comparison between correct positive and negative responses. In order to correctly reject a test stimulus, each element in the memory set must be compared with this item (1). Thus, the RT for a negative response should be determined by the duration of the total number of comparisons in the memory set. A correct positive response can be made, presumably, whenever a match is detected between the test stimulus and its representation in memory. Such a match will, on average, be detected halfway through the serial scan. If the serial comparison process terminates when a match is detected, then the slope for positive responses should be half the slope for negative responses. This is because, on the average, there will be half as many comparisons for positive responses as for negative responses before a match is detected. The slopes that Sternberg observed were, in fact, approximately equal suggesting that the same number of comparisons are made for both positive and negative stimuli. This implies that, regardless of response type, all comparisons between the test stimulus and elements in the memory

set are made prior to the binary decision. Sternberg concluded that the memory search is an exhaustive, serial comparison, process.

Several investigators have examined the event related brain potentials (ERP)s elicited in a Sternberg task in an effort to provide an assessment of memory search time uncontaminated by motor response processes. The P300 is a component of the ERP which is a manifestation of intracranial activity involved in cognitive processing (Donchin 1979, 1981). The P300 is thought to be elicited when the current mental model (schema) of the environment does not match environmental events. This is consistent with the view that the P300 "subroutine" performs tasks that are required in the maintenance (or context updating) of working memory (Donchin 1979, 1981; Donchin and Bashore, in press). Evidence supporting the context updating hypothesis about the functional significance of the P300 was reported by Karis, Fabiani, and Donchin (1984). These investigators reported that, for subjects who employed rote strategies, P300 amplitude predicted <u>subsequent</u> recall and recognition performance in a memory task. The larger the P300 amplitude, the greater the probability that an item was recognized or recalled or both. The latency of the P300 component reflects the relative time to evaluate and categorize a stimulus, but is relatively insensitive to processes related to response selection and execution (Kutas, McCarthy, and Donchin, 1977; McCarthy and Donchin, 1981; Magliero, Bashore, Coles, and Donchin, 1984).

Studies examining the P300 latency in the Sternberg task (Roth, Kopell, Tinklenberg, Darley, Sikara, and Vesecky, 1975; Gomer, Spicuzza, and O'Donnell, 1976; Roth, Tinklenberg, and Kopell, 1977; Adam and Collins,

1978; Ford, Roth, Mohs, Hopkins, and Kopell, 1979; Pfefferbaum, Ford, Roth, and Kopell, 1980; Ford, Pfefferbaum, Tinklenberg, and Kopell, 1982) have generally reported parallel RT regression lines (i.e., equal slopes) for positive and negative stimuli, replicating the results obtained by Sternberg. Furthermore, the P300 latency slopes for positive and negative stimuli were also equal, but the slopes for P300 latency were approximately half the slopes of the RT data. Ford et al., (1979) interpreted the slope of P300 latency to reflect the time necessary to evaluate the test stimulus against each element in the memory set, and suggested that P300 latency slope is a better measure of stimulus evaluation time than the RT, because the RT slope may also include response related processes which increase in duration as a function of set size.

The previous ERP Sternberg studies employed a procedure in which the memory set was changed from trial to trial. In this procedure there is no benefit in updating working memory when a test stimulus is presented because a new memory set is presented prior to each trial. Therefore, it is unclear what role, consequence, or functional significance the P300 plays in these earlier studies The present experiment employs a procedure in which the memory set is held constant for a block of trials. In this procedure there is a benefit in updating working memory when a test stimulus is presented. Theios (Theios, Smith, Haviland, Traupmann, and Moy, 1973; Theios and Walter,1974), employing a procedure in which the memory set was held constant for a series of trials, reported that the response was speeded when stimuli were repeated within a short time interval. These sequential effects suggest that the memory representations are rearranged on a trial by trial

basis so as to establish or maintain the current representation of the frequent stimuli in working memory for use on subsequent trials. If the P300 reflects the updating of working memory, then the P300 should reflect this trial by trial restructuring. The present experiment evaluates the role of the P300 when there is a benefit in updating working memory after each trial.

Methods

Subjects

Twenty-five right handed male students (age 18 - 25) from the University of Illinois served as subjects. They were paid $4.50 for participating in the experiment.

Stimuli

Stimuli were presented on an Imlac display. The set of stimuli was comprised of 19 of the letters in the alphabet (excluding the vowels and the letters Q and V). The stimuli were 19mm in height and 11mm in width. The visual angle subtended 1.21 degrees.

Procedure

Subjects were seated in front of the display. The experiment began by presenting 1, 2, 3, 4, or 5 letters for subjects to memorize. This defined the memory set. Elements in the memory set were presented sequentially in the center of the screen for 200 msec with an interstimulus interval of 2000 msec. A 500 msec, 800 Hz, warning tone signaled the end of the memory set. Following a 1000 msec delay, a series of 30 letters (the test stimuli) were each presented for 200 msec, with an interstimulus interval of 2000 msec. Subjects were instructed to press one of two buttons to indicate

whether or not the test stimulus was a member of the memory set. Buttons were counterbalanced across subjects. Subjects were also instructed to respond as rapidly as possible without making errors. Fifteen of the test stimuli were members of the memory set (positive stimuli) and the other fifteen were not (negative stimuli). For a given memory set, each member served as a test stimulus the same number of times (except for set sizes 2 and 4 where one of the memory set elements was presented one less time than the other elements in the memory set). The negative stimuli were drawn randomly from a subset of the letters not included in the memory set (2). The number of elements in the subset remained constant (14 elements) across set size.

The experiment was organized in 23 blocks. Each block consisted of a memory set followed by 30 test stimulus presentations. The first three blocks provided practice. The practice blocks consisted of a set size of 1, a set size of 3, and a set size of 5. Then twenty blocks were randomly presented, four blocks (120 trials) for each set size. An interblock interval of 13.1 seconds was employed. During the interblock interval the word "RELAX" was presented on the display. 2000 msec prior to the end of the interblock interval a 500 msec, 800 Hz, tone was presented to signal the upcoming memory set. Three rest periods were provided during the session to prevent fatigue. During the rest interval the word "REST" was presented on the display. Subjects were allowed to rest for a duration ranging from 2 to 5 minutes. Approximate running time was 45 minutes.

Data Collection

Burden Ag/AgCl electrodes were affixed with collodion at Fz, Cz, and Pz

Vesecky, T. B. (1975). The contingent negative variation during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 38, 171-174.

Roth, W. T., Tinklenberg, J. R., & Kopell, B. S. (1977). Ethanol and marihuana effects on event-related brain potentials in a memory retrieval paradigm. Electroencephalography and Clinical Neurophysiology, 42, 381-388.

Sternberg, S. (1966). High-speed scanning in human memory. Science, 153, 652-654.

Sternberg, S. (1967). Two operations in character recognition: Some evidence from reaction-time measurements. Perception and Psychophysics, 2, 45-53.

Sternberg, S. (1969a). Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 57, 421-457.

Sternberg, S. (1969b). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), Attention and Performance II. Acta Psychologica, 30, (pp. 276-315).

Sternberg, S. (1975). Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 27, 1-32.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. (1973). Memory scanning as a serial self-terminating process. Journal of Experimental Psychology, 97, 323-336.

Theios, J., & Walter, D. G. (1974). Stimulus and response frequency and sequential effects in memory scanning reaction times. Journal of Experimental Psychology, 102, 1092-1099.

Physiological Psychology, 4, 61-65.

Gratton, G., Coles, M. G. H., Donchin, E. (1983a) A new method for off-line removal of ocular artifact. Electroencephalography and Clinical Neurophysiology, 55, 468-484.

Gratton, G., Coles, M. G. H., Donchin, E. (1983b) Filtering for spatial distribution: A new approach (vector filter). Psychophysiology, 20, 443-444. (Abstract)

Jasper, H. H. (1958). The ten twenty electrode system of the international federation. Electroencephalography and Clinical Neurophysiology, 10, 371-375.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and memory: Individual differences in the von Restorf effect. Cognitive Psychology, 16, 177-216.

Kutas, M., McCarthy, G., Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792-795.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin E. (1984). On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21, 171-186.

Mandler, G. (1980). Recognizing: The judgement of previous occurrence. Psychological Review, 87, 252-271.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. Science, 211, 77-80.

Pfefferbaum, A., Ford, J. M., Roth, W. T., & Kopell, B. S. (1980). Age differences in P3-reaction time associations. Electroencephalography and Clinical Neurophysiology, 49, 257-265.

Roth, W. T., Kopell, J. R., Tinklenberg, C. F., Darley, Sikora, R., &

References

Adam, N., & Collins, G. I. (1978). Late components of the visual evoked potential to search in short-term memory. Electroencephalography and Clinical Neurophysiology, 44, 147-156.

Coles, M. G. H., Gratton, G., Kramer, A. F., & Miller, G. A. (in press). Principles of signal acquisition and analysis. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, processes, and applications. Vol I: Systems. New York: Guilford Press.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human processing. In H. Begleiter (Ed.), Evoked potentials and behavior, (pp. 13-75). New York: Plenum.

Donchin, E. (1981). Surprise!...surprise? Psychophysiology, 18, 493-513.

Donchin, E., & Bashore, T. R. (in press). Clinical versus psychophysiological paradigms in the study of event related brain potentials. The Behavioral and Brain Sciences.

Ford, J. M., Pfefferbaum, A., Tinklenberg, J. R., & Kopell, B. S. (1982). Effects of perceptual and cognitive difficulty on P3 and RT in young and old adults. Electroencephalography and Clinical Neurophysiology, 54, 311-321.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, III, W. F., & Kopell, B. S. (1979). Event-related potentials recorded from young and old subjects during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 47, 450-459.

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. (1976). Evoked potential correlates of visual item recognition during memory-scanning tasks.

to the mechanisms involved. One possible route is a process of retrieval from long-term memory through familiarity and identification (Mandler, 1980). Another possible route is a rapid exhaustive scan (either serial or parallel) through all of the letters and their S-R relationships, which are stored in long-term memory. According to these two views, the process, which has a relatively constant latency, is initiated when the test stimulus is encoded and procedes in parallel with the serial, exhaustive search of working memory. If a match is detected in working memory the process is terminated. If a match was not detected in working memory, the process continues and a P300 is elicited when the information is scrolled into working memory. It remains for future research to resolve this issue.

The goal of this experiment was to investigate the relationship between RT and P300 latency when there is a benefit to update working memory to further understand the cognitive processes involved in the Sternberg task. The data suggest that the P300 is associated with a process which is elicited to maintain or establish the representation of the test stimulus in working memory for use on subsequent trials.

related to performance on current trials.

The above argument suggests that there will be sequential effects for negative stimuli. If a negative stimulus is not contained in working memory the response will be based on the absence of a match, resulting in a longer RT. If the same stimulus is presented on a subsequent trial it will be contained in working memory. The response will not be delayed, as in the case of the stimuli which are not contained in working memory, therefore the RT and the RT-P3 coupling should be similar to that of the positive stimuli. Recall that Theios et al. (1974) reported sequential effects for negative stimuli. Unfortunately in the present experiment, there were not enough trials to adequately examine the sequential effects suggested by this model.

These results are quite different from earlier ERP Sternberg studies in which the memory set was changed prior to each trial. One explanation for the discrepant findings is the role context updating plays when more than one trial follows each memory set. One avenue of future research will be to manipulate, on a within subject basis, the number of test stimuli following the memory set. If the degree of context updating is affected, then P300 latency should reflect this manipulation. Another area of interest will be to examine the sequential effects which are predicted for negative stimuli. If negative stimuli are repeated within a short interval, RT should be speeded and P300 latency should be similar to the P300s elicited by positive stimuli.

How the information is retrieved and subsequently scrolled into working memory remains to be elucidated. The process has a relatively constant latency (i.e., one which is not affected by set size) which provides a clue

of the P300 acting to update the current mental model of the environment (in this case the letters compared with the test stimulus) for future trials. The sequential effects for negative stimuli reported by Theios et al. (1974) suggest that some of the recent negative stimuli may be stored in working memory. The P300 may be a manifestation of the process of scrolling the negative stimulus into working memory for use on subsequent trials.

P300 amplitude may be larger for positive stimuli for several reasons. The detection of a matching element in working memory may elicit a larger P300 than if no match was detected. Further, the subject is actively involved in keeping track of the task relevant stimuli. The P300 can be seen as an active process, strengthening the S-R relationships of the stimuli in working memory for future trials.

These data suggest that the response is a function of the same processes for both positive and negative stimuli, but negative stimuli have an additional delay, perhaps because the response is based on the absence of a match. The P300 process for positive stimuli is related to the RT process, however the P300 is a manifestation of context updating of working memory for subsequent trials. The P300 process for negative stimuli is future oriented and not related to performance on current trials. The benefit of updating the elements stored in working memory is that if the test stimulus which was not contained in working memory is repeated within a short time interval it will reside in working memory and the response will not have to be made on the basis of the absence of a match. Thus the P300 is elicited to maintain or establish the representation of the stimulus in working memory for future trials. This context updating is not directly

accomplished in the same fashion, namely an exhaustive, serial comparison process, and that the number of comparisons made in working memory covaries with set size. The difference in the intercepts may be the result of responding in the absence of the detection of a match between the negative stimulus and the elements in the memory set.

For positive stimuli, the processes which are manifested by the P300 are related to the processes underlying the RT data, but this is not the case for negative stimuli. There is a decoupling between RT and P300 latency for negative stimuli. This decoupling takes the form of no set size effect for P300 latency. In addition, P300 amplitude is smaller for negative stimuli than for positive stimuli. Thus the P300 data suggest that positive and negative trials differ in some underlying processes.

Set size 1 appears to be quite different from set sizes 2-5. The mean RT for set size 1 for both positive and negative stimuli falls below the regression line. Further, the P300s elicited by the negative stimuli in set size 1 occur earlier than those elicited by set sizes 2-5. The above argument suggests that the processes involved in set size 1 are not the same as those involved in set sizes 2-5. This makes intuitive sense if you consider that, for set size 1, the task is similar to a same-different judgement. The serial comparison process need not be invoked until there are 2 or more elements in the memory set.

The P300 is not affected by set size for negative stimuli while RT is affected by set size. This is consistent with the suggestion that the P300 is not necessary for performance on current trials, but instead may be related to performance on future trials. This is consonant with the notion

Discussion

The major findings of this study were: 1) A replication of Sternberg's findings with the RT data. RT increased as a function of set size for both positive and negative stimuli, and the slopes of the two regression lines were approximately equal. There was an additional amount of time taken for negative stimuli, which resulted in a larger intercept for those stimuli. 2) P300 latency increased as a function of set size for positive stimuli, but not for negative stimuli. For negative stimuli, P300 latency was approximately fixed for set sizes 2-5, while the P300 for set size 1 occurred earlier than for the larger set sizes. 3) There was a stronger correlation between RT and P300 latency for positive stimuli than for negative stimuli. This relationship existed both within and across set size. 4) P300 amplitude was larger for positive stimuli than for negative stimuli. There were no P300 amplitude differences as a function of set size.

In sum, for positive stimuli both RT and P300 latency increased as a function of set size. RT increased as a function of set size for negative stimuli, but P300 latency was relatively fixed for set sizes 2-5. This dissociation between RT and P300 latency provides a clue to the underlying processes involved in the Sternberg task.

The parallel slopes of the RT data suggest that the processes which underlie the <u>response</u> are the same for both positive and negative stimuli, but there is an additional amount of time taken for negative stimuli. Since RT increases as a function of set size equivalently for positive and negative stimuli, it implies that the scanning of the memory set is

---

Insert Figure 6 About Here

---

2 was labelled "P300" based on the latency and scalp distribution. The latency adjustment procedure blends the other components in the ERP making interpretation of components other than P300 difficult, therefore only component 2 was analyzed. A 3 factor (response type x set size x electrode) repeated measures analysis of variance was performed on the PCA component scores to test for differences in P300 amplitude across conditions. The P300s elicited by positive stimuli were larger (component score=-.094) than for negative stimuli (component score=.168), $F(1,24)=50.92$. Figure 7 presents the latency adjusted Pz electrode overplots of positive and negative stimuli for set sizes 1-5. P300 amplitude did not vary as a function of set size (component scores: set size 1=.043, set size 2=.085, set size 3=.010, set size 4=.019, set size 5=.028), $F(4,96)=0.17$. Figure 8 presents the latency adjusted Pz electrode overplots of set size for positive and negative stimuli. The difference in P300 amplitude between

---

Insert Figures 7 & 8 About Here

---

positive and negative stimuli remained constant over set size, $F(4,96)=1.42$. The only other significant result was a response type x electrode interaction, $F(4,96)=92.66$. This was due to a divergence at the Cz electrode when collapsing across set size for the two response types.

negative stimuli at each set size. The difference between the Z transformed correlations for positive and negative stimuli was not significant for set size 1, $t(24)=.25$, nor for set size 2, $t(24)=1.48$, but was significant for set size 3, $t(24)=3.25$, for set size 4, $t(24)=3.96$, and for set size 5, $t(24)=5.32$. These Z transformed correlations were entered in a 2 factor (response type x set size) repeated measures analysis of variance. The

---

Insert Table 1 About Here

---

correlations were larger for positive stimuli than for negative stimuli, $F(1,24)=34.67$. The correlations did not vary as a function of set size when pooled over response type, $F(4,96)=0.60$; however, the correlations for positive stimuli increased as a function of set size while for negative stimuli the correlations diminished as a function of set size, $F(4,96)=4.46$.

P300 amplitude

To control for latency jitter, single trials were latency adjusted by shifting the P300 peak as defined in the P300 latency analysis (see above) to a common point (500 msec). Subject averages were calculated for each condition. The subject averages were then submitted to a principal component analysis (PCA) involving 750 waveforms (25 subjects x 2 response types x 5 set sizes x 3 electrodes). Three components were rotated using a varimax procedure. The component loadings are shown in figure 6. Component

stimuli. P300 latency increased as a function of set size at a greater rate

---

Insert Figures 3,4,& 5 About Here

---

for positive stimuli than for negative stimuli, $F(4,96)=7.82$. The P300 latency regression equations for positive and negative stimuli were:

$$\text{Positive stimuli:} \quad P3' = 18.0\,(X) + 502.0$$

$$\text{Negative stimuli:} \quad P3' = 7.0\,(X) + 581.5$$

A 2 factor (response type x set size) repeated measures analysis of variance was performed on the standard deviation of the P300 latency data. The P300s elicited by negative stimuli (165.5) varied more than for positive stimuli (150.6), $F(1,24)=19.19$. The standard deviation of P300 latency did not vary as a function of set size, (set size 1=162.4, set size 2=153.5, set size 3=155.3, set size 4=158.6, set size 5=160.6), $F(4,96)=1.99$. The difference in the standard deviation of P300 latency for positive and negative stimuli remained constant across set size, $F(4,96)=1.48$.

To further explore the relationship between RT and P300 latency, correlations based on single trial estimates of P300 latency and RT were calculated. For every subject, the correlation for positive stimuli was larger than the correlation for negative stimuli. The mean correlation for positive stimuli was $r=.25$. The mean correlation for negative stimuli was $r=.09$. The difference between the Z transformed correlations was significant, $t(24)=8.30$. Correlations between RT and P300 latency were also computed for positive and negative stimuli within each set size. Table 1 presents the mean correlations between RT and P300 latency for positive and

Event Related Potential Data

Figure 2 presents the superaverage waveforms at the Fz, Cz, and Pz electrode sites for the 2 (response type) x 5 (set size) conditions. The predominant component in these ERP data is the P300 - a positive potential which occurs at least 300 msec after stimulus onset and is parietally maximal.

Insert Figure 2 About Here

P300 latency

A single trial estimate of P300 latency was employed in all analyses involving ERPs. Single trial P300 latency was determined by forming a composite waveform with a vector filter (Gratton, Coles, Donchin, 1983b; Coles, Gratton, Kramer, & Miller, in press) and obtaining the maximum cross correlation with the positive segment of a .5 Hz sine wave. Trials with maximum cross correlations below r=.3 were excluded from analyses involving ERPs.

A 2 factor (response type x set size) repeated measures analysis of variance was performed on the P300 latency data. Figure 3 presents the mean P300 latencies for positive and negative stimuli for set sizes 1-5. The P300s elicited by positive stimuli (556.0) occurred earlier than for negative stimuli (604.2), $F(1,24)=146.83$. Figure 4 presents the Pz electrode overplots of response type for set sizes 1-5. Furthermore, P300 latency increased as a function of set size, (set size 1=552.3, set size 2=564.3, set size 3=585.2, set size 4=600.5, set size 5=598.2), $F(4,96)=24.15$. Figure 5 presents the Pz electrode overplots of set size for positive and negative

----

Insert Figure 1 About Here

----

stimuli remained constant across set size, $F(4,96)=2.12$. The RT regression equations for positive and negative stimuli were:

$$\text{Positive stimuli:} \quad RT' = 45.1\ (X) + 428.5$$

$$\text{Negative stimuli:} \quad RT' = 40.0\ (X) + 504.0$$

A 2 factor (response type x set size) repeated measures analysis of variance was performed on the standard deviation of the RT data. Subjects' responses were more variable for negative stimuli (147.1) than for positive stimuli (135.9), $F(1,24)=12.53$. The standard deviation of RT increased as a function of set size, (set size 1=122.8, set size 2=128.5, set size 3=138.8, set size 4=156.2, set size 5=161.3), $F(4,96)=17.92$. The difference between the standard deviation of RT for positive and negative stimuli remained constant across set size, $F(4,96)=2.80$.

A 2 factor (response type x set size) repeated measures analysis of variance was performed on the error data. An error was defined as a trial in which the subject made an incorrect response (3). Subjects made more errors for positive stimuli (3.5%) than for negative stimuli (1.1%), $F(1,24)=38.07$. Error rates increased as a function of set size, (set size 1=1.0%, set size 2=1.6%, set size 3=2.8%, set size 4=3.0%, set size 5=3.1%), $F(4,96)=8.36$; however, error rates increased at a greater rate for positive stimuli than for negative stimuli, $F(4,96)=5.58$.

(10/20 system; Jasper, 1958) and with stomaseal adhesive collars to the reference sites (linked mastoids), the ground (forehead), and the EOG sites (sub- and supra-orbital). Electrode impedance did not exceed 10 KOhms. The EEG signals were amplified with a Van Gogh model 50000 amplifier using a 10 second time constant and an upper half-amplitude frequency 35 Hz, 3 dB/octave roll off filter. The EEG signals were digitized at the rate of 250 Hz for 1600 msec, beginning 100 msec prior to stimulus onset. All aspects of experimental control and data collection were controlled by a DEC PDP-11/40 computer interfaced with an Imlac graphics processor. Data was collected whenever a letter was displayed and stored on magnetic tape for subsequent quantification and analysis. Eye movement artifacts were corrected off-line using a procedure described by Gratton, Coles, and Donchin (1983a).

Results

For all analyses, only trials in which the subject's response was correct were examined. A significance level of .01 was adopted for all inferential tests.

Reaction time

A 2 factor (response type x set size) repeated measures analysis of variance was performed on the RT data. Figure 1 presents the mean RT for positive and negative stimuli at set sizes 1-5. Subjects responded faster to positive stimuli (563.7) than to negative stimuli (624.0), $F(1,24)=65.68$. Furthermore, the larger the set size the slower the RT, (set size 1=495.4, set size 2=559.1, set size 3=603.4, set size 4=647.5, set size 5=664.0), $F(4,96)=156.82$. The difference between the RTs for positive and negative

Footnotes

(1)Sternberg's model assumes that only the elements in the memory set are compared with the test stimulus. Further, his model assumes that elements in the memory set are sampled randomly, without replacement during the serial comparison process.

(2)A restriction on the random sampling was imposed so that no particular negative stimulus was presented consecutively.

(3)Trials in which subjects failed to make a response within the 1500 msec time limit occurred less than 1% of the time.

Table 1

Reaction time X P300 latency correlations within each set size

	1	2	3	4	5
Pos	.119	.207	.244	.245	.250
Neg	.129	.119	.092	.060	-.006

Figure Captions

Figure 1. Mean reaction time for positive (solid line) and negative (dashed line) stimuli for set sizes 1-5. Responses were faster for positive stimuli than negative stimuli. Reaction time increased at the same rate for positive and negative stimuli across set size.

Figure 2. Electrode distribution of the superaverage waveforms for the 2 (response type) x 5 (set size) conditions. Fz is represented by solid lines, Cz is represented by dashed lines, and Pz is represented by dotted lines. The predominant component in these waveforms is the P300 - a positive potential which occurs at least 300 msec after stimulus onset and is parietally (Pz) maximal.

Figure 3. Mean P300 latency for positive (solid line) and negative (dashed line) stimuli for set sizes 1-5. P300 latency occurred earlier for positive than negative stimuli. P300 latency increased as a function of set size; however, P300 latency increased at a greater rate for positive stimuli than negative stimuli.

Figure 4. Overplot of Pz superaverage waveforms for positive (solid lines) and negative (dashed lines) stimuli for set sizes 1-5. P300 latency occurred earlier for positive than negative stimuli.

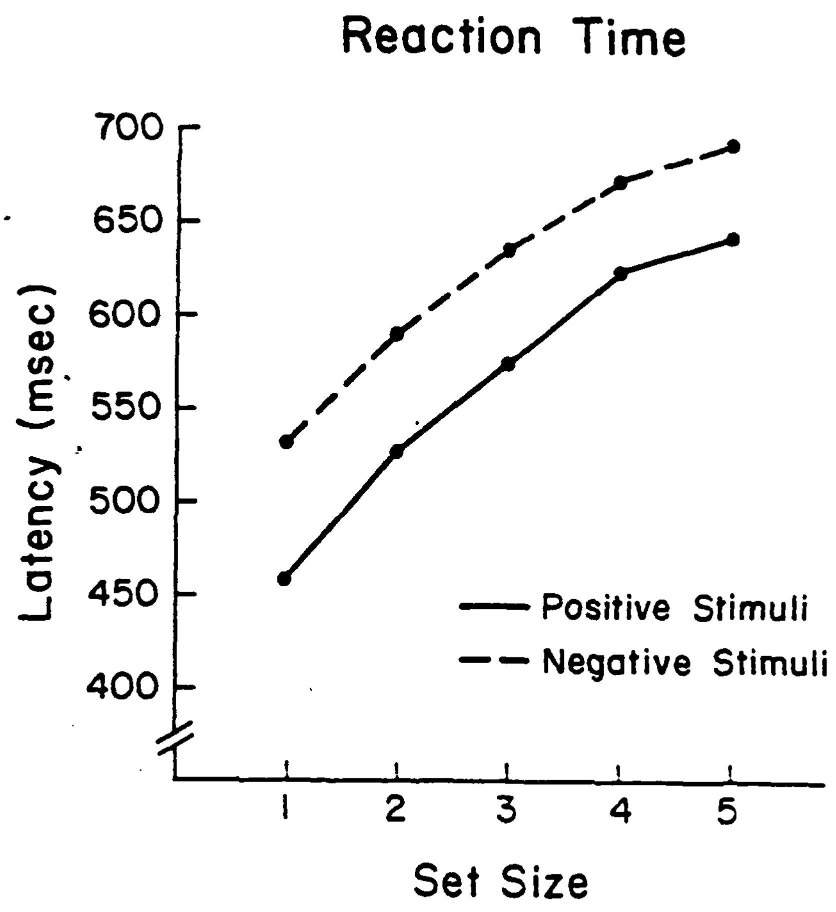
Figure 5. Overplot of Pz superaverage waveforms for set sizes 1-5 for positive and negative stimuli. P300 latency increases as a function of set size for positive stimuli, but for negative stimuli P300 latency was relatively fixed for set sizes 2-5.

Figure 6. Grand mean waveform and component loadings for the three components extracted from the principal components analysis. The solid line

represents the grand mean waveform. The dotted line represents +1 standard deviation from the grand mean waveform and the dashed line represents -1 standard deviation from the grand mean waveform. Examination of the component loadings reveals that the second component extracted in the principal component analysis loads on the P300 component.

Figure 7. Overplot of latency adjusted Pz superaverage waveforms for positive (solid lines) and negative (dashed lines) stimuli for set sizes 1-5. P300 amplitude was larger for positive stimuli than for negative stimuli.

Figure 8. Overplot of latency adjusted Pz superaverage waveforms for set sizes 1-5 for positive and negative stimuli. P300 amplitude does not vary as a function of set size.



Reaction Time
700
650
600
550
500
450
400
Latency (msec)
Positive Stimuli
Negative Stimuli
1
2
3
4
5
Set Size

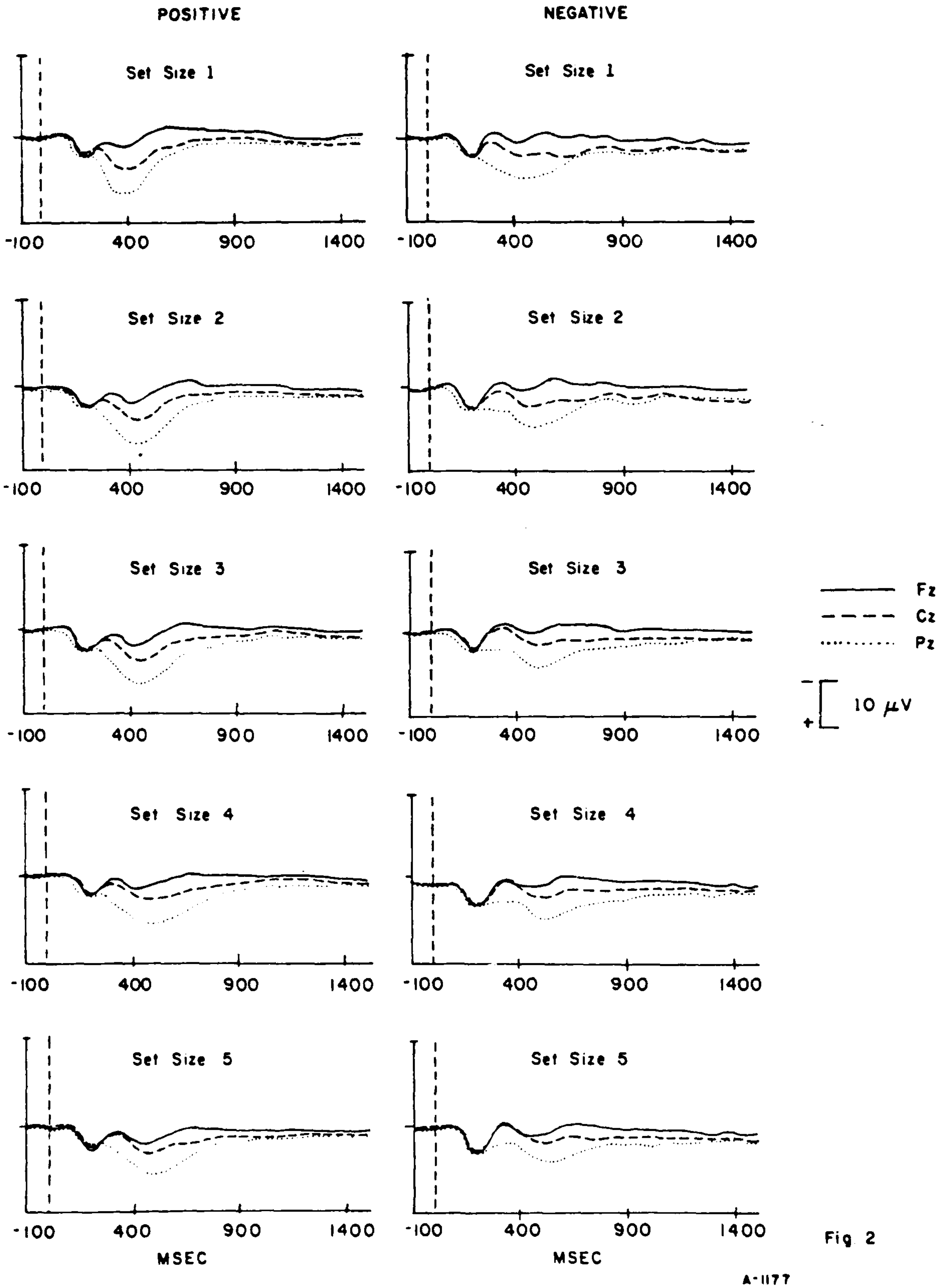
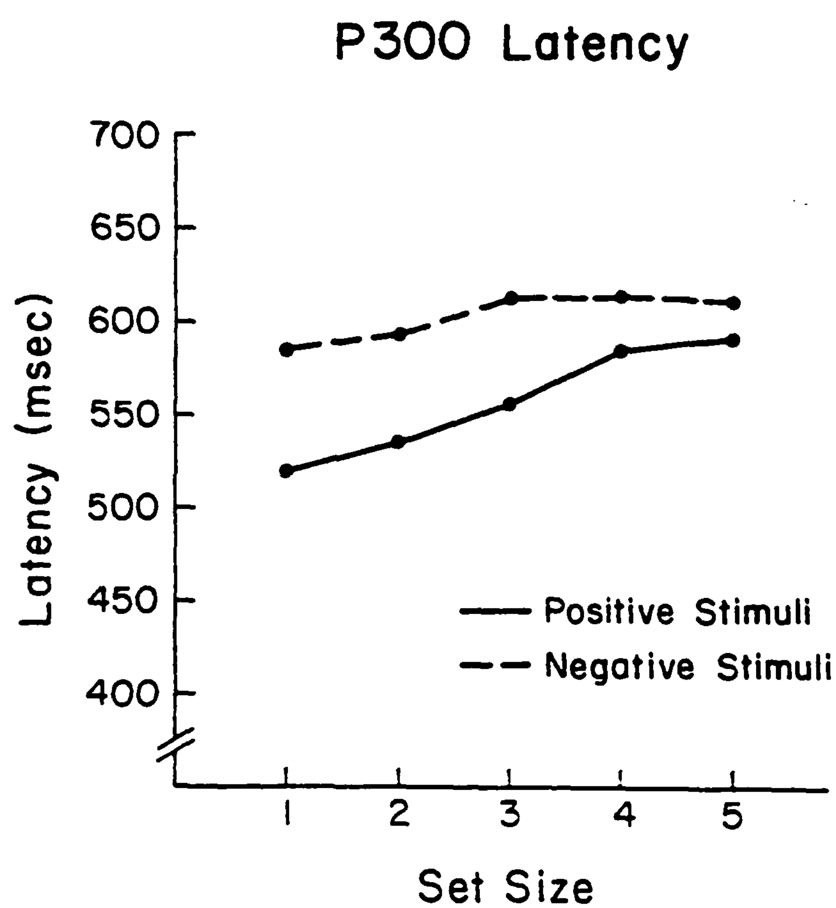
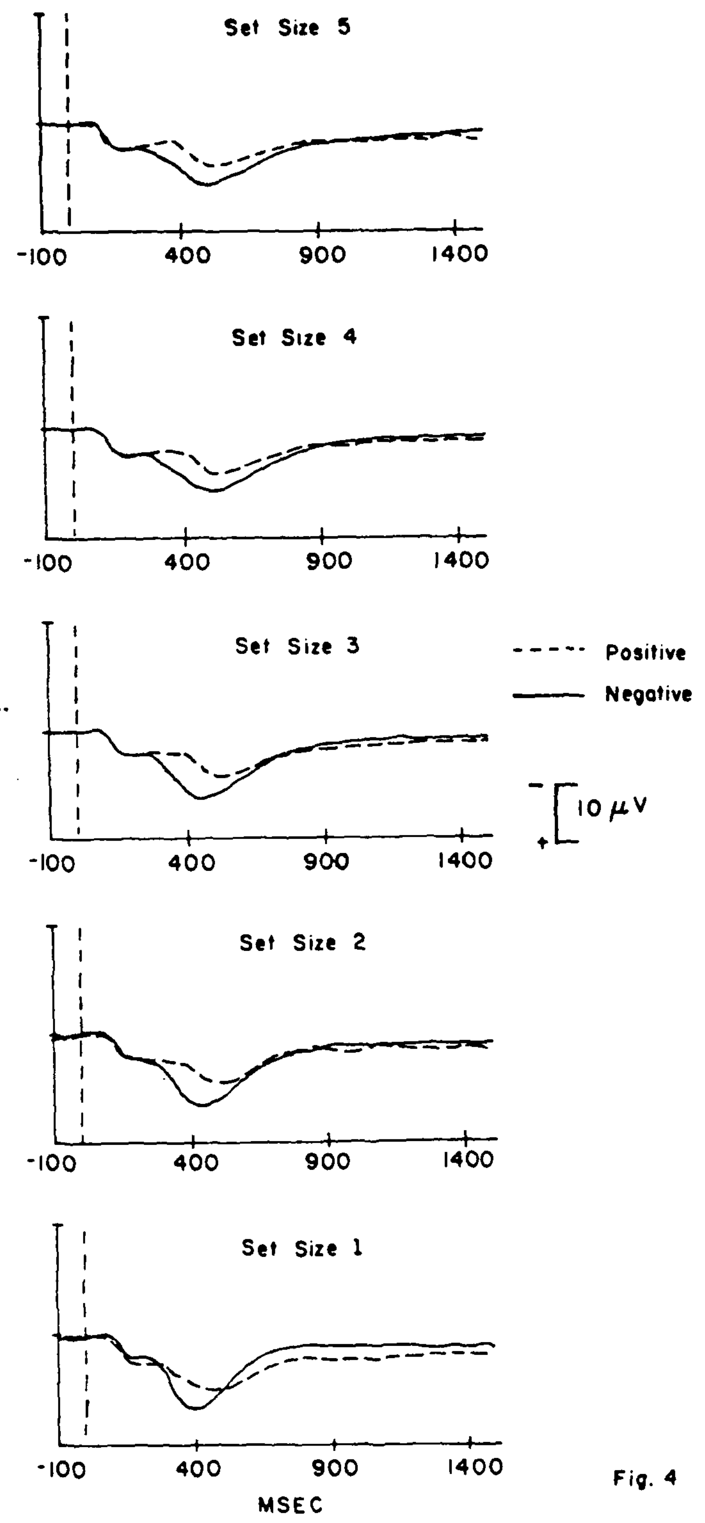

POSITIVE
NEGATIVE
Set Size 1
Set Size 2
Set Size 3
Set Size 4
Set Size 5
-100
400
900
1400
MSEC
Fz
Cz
Pz
10 μV
A-1177

Fig 2

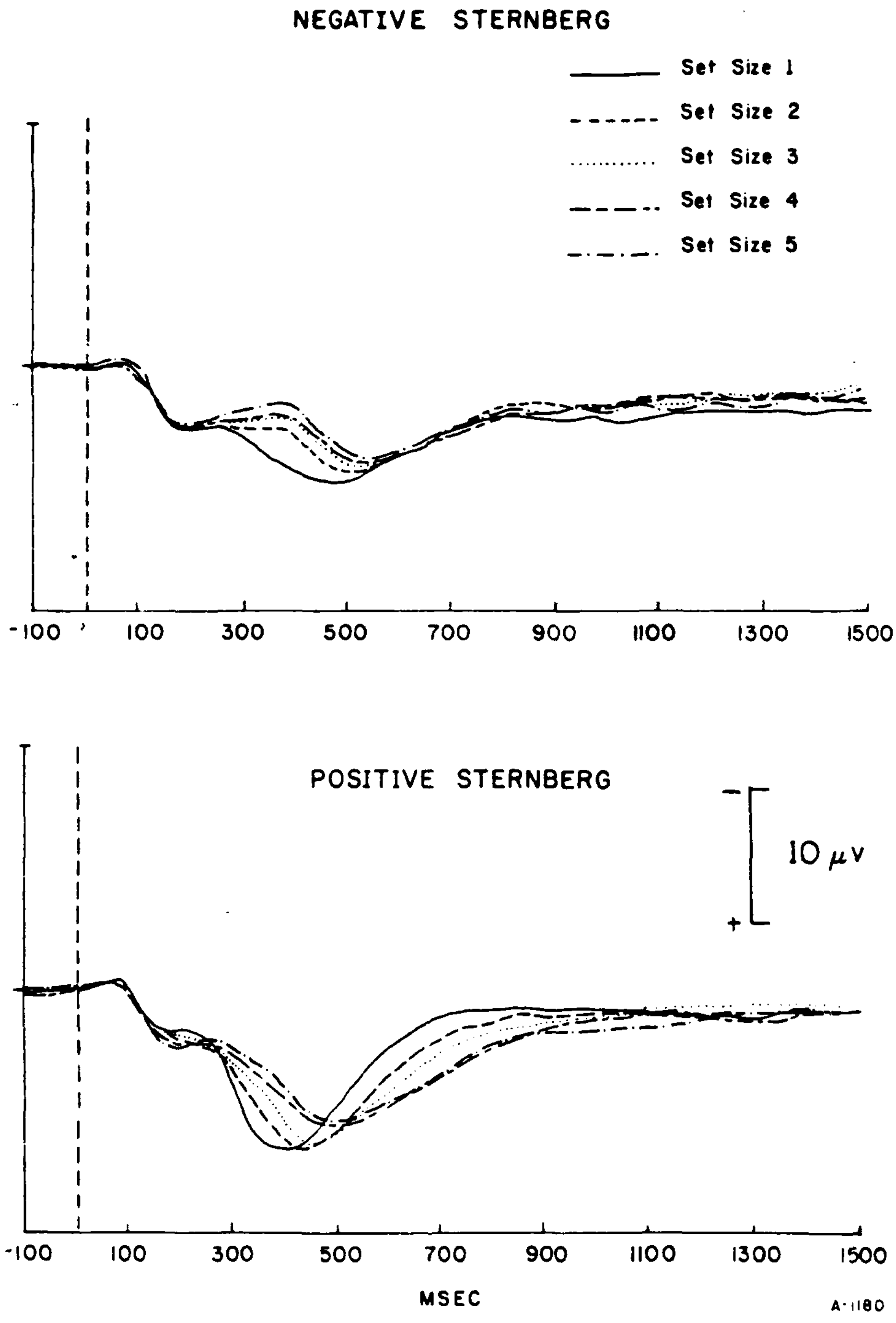

P300 Latency
Latency (msec)
700
650
600
550
500
450
400
1
2
3
4
5
Set Size
Positive Stimuli
Negative Stimuli



Set Size 5
-100
400
900
1400
Set Size 4
-100
400
900
1400
Set Size 3
-100
400
900
1400
Positive
Negative
10 μV
Set Size 2
-100
400
900
1400
Set Size 1
-100
400
900
1400
MSEC

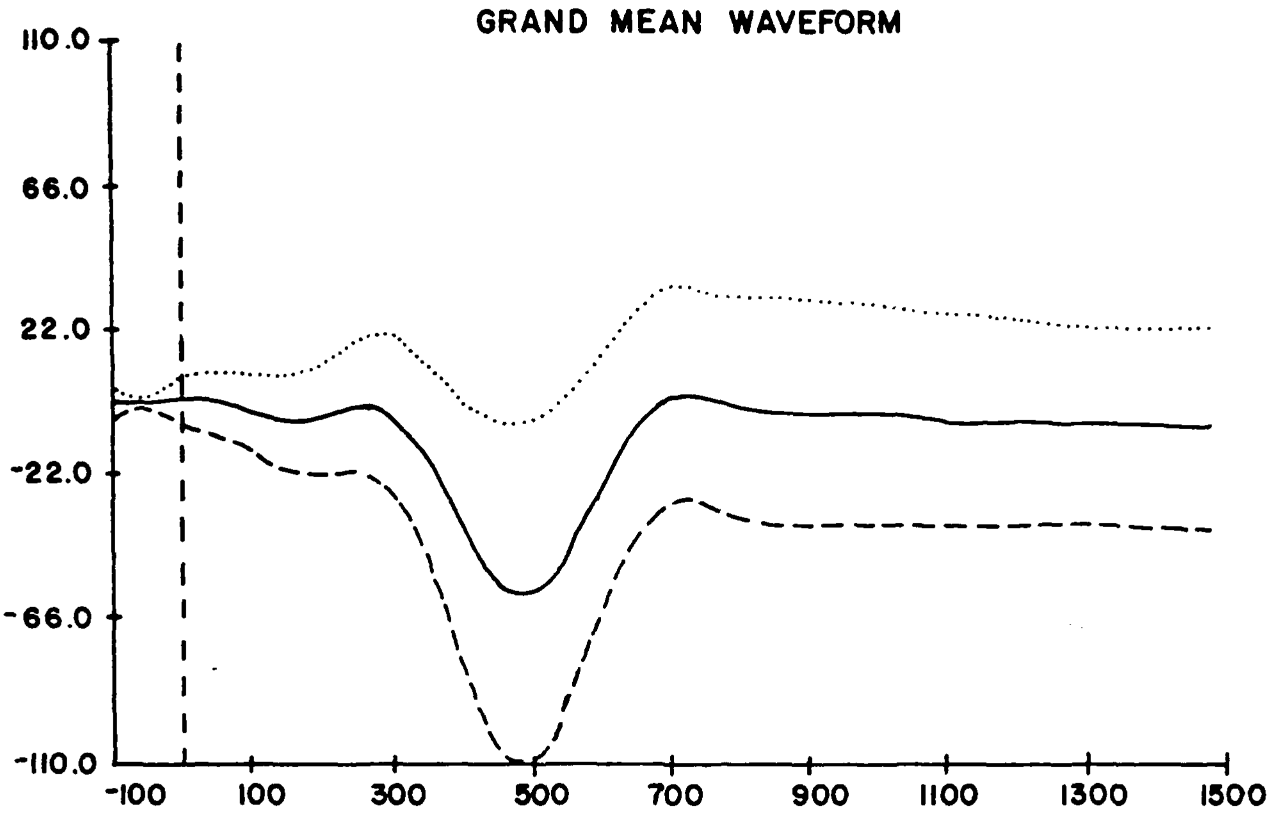
Fig. 4

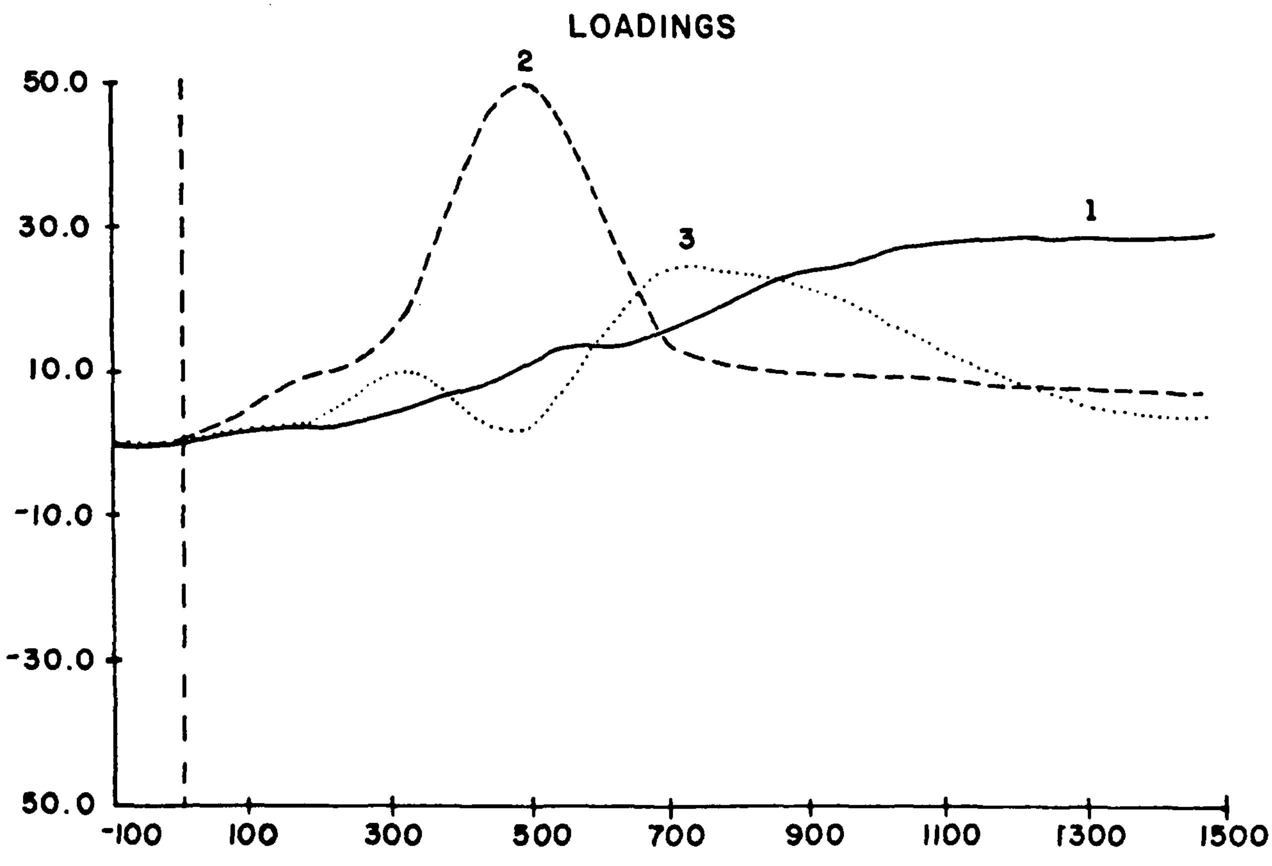
A-1178



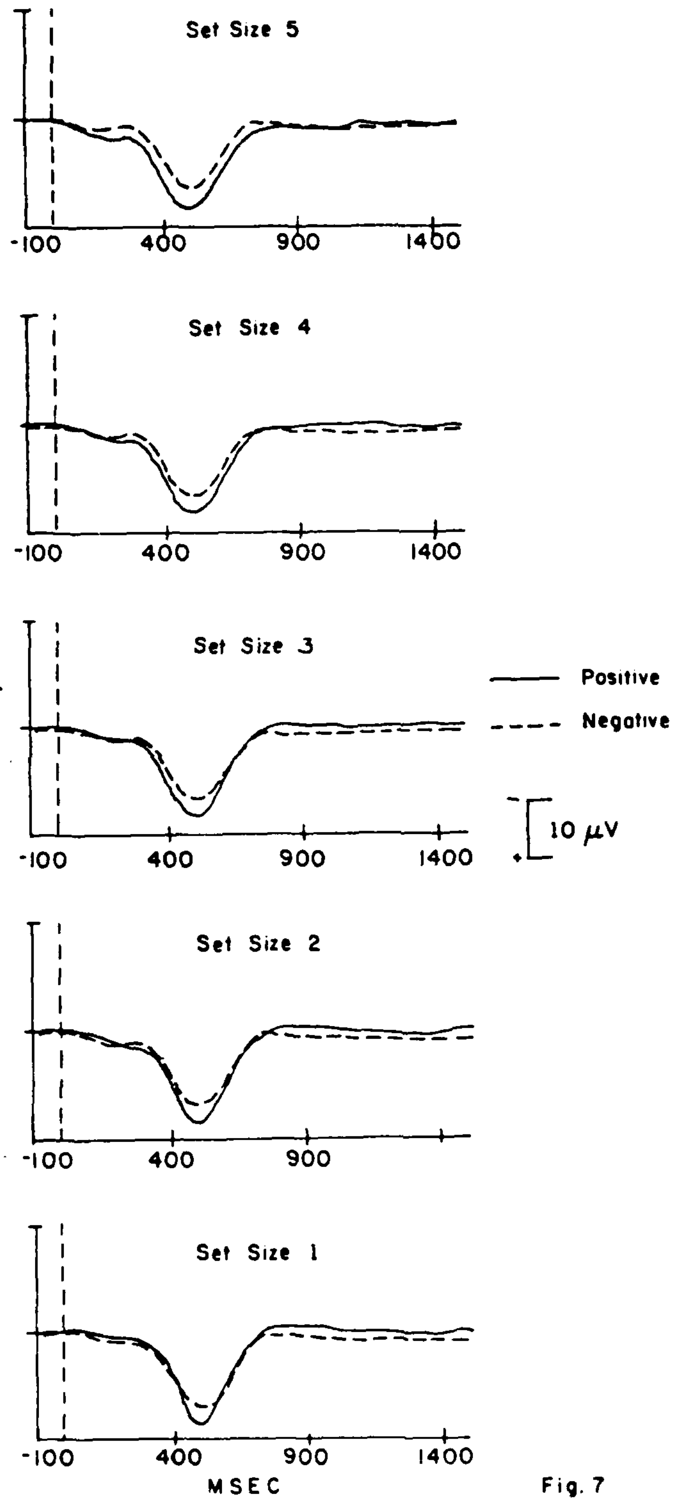
NEGATIVE STERNBERG
Set Size 1
Set Size 2
Set Size 3
Set Size 4
Set Size 5
-100
100
300
500
700
900
1100
1300
1500
POSITIVE STERNBERG
10 μv
-100
100
300
500
700
900
1100
1300
1500
MSEC
A-1180

Fig. 5


GRAND MEAN WAVEFORM
110.0
66.0
22.0
-22.0
-66.0
-110.0
-100
100
300
500
700
900
1100
1300
1500


LOADINGS
2
3
1
50.0
30.0
10.0
-10.0
-30.0
50.0
-100
100
300
500
700
900
1100
1300
1500

A-1181



Set Size 5
-100
400
900
1400
Set Size 4
-100
400
900
1400
Set Size 3
Positive
Negative
10 μV
-100
400
900
1400
Set Size 2
-100
400
900
Set Size 1
-100
400
900
1400
MSEC

Fig. 7

A-1179

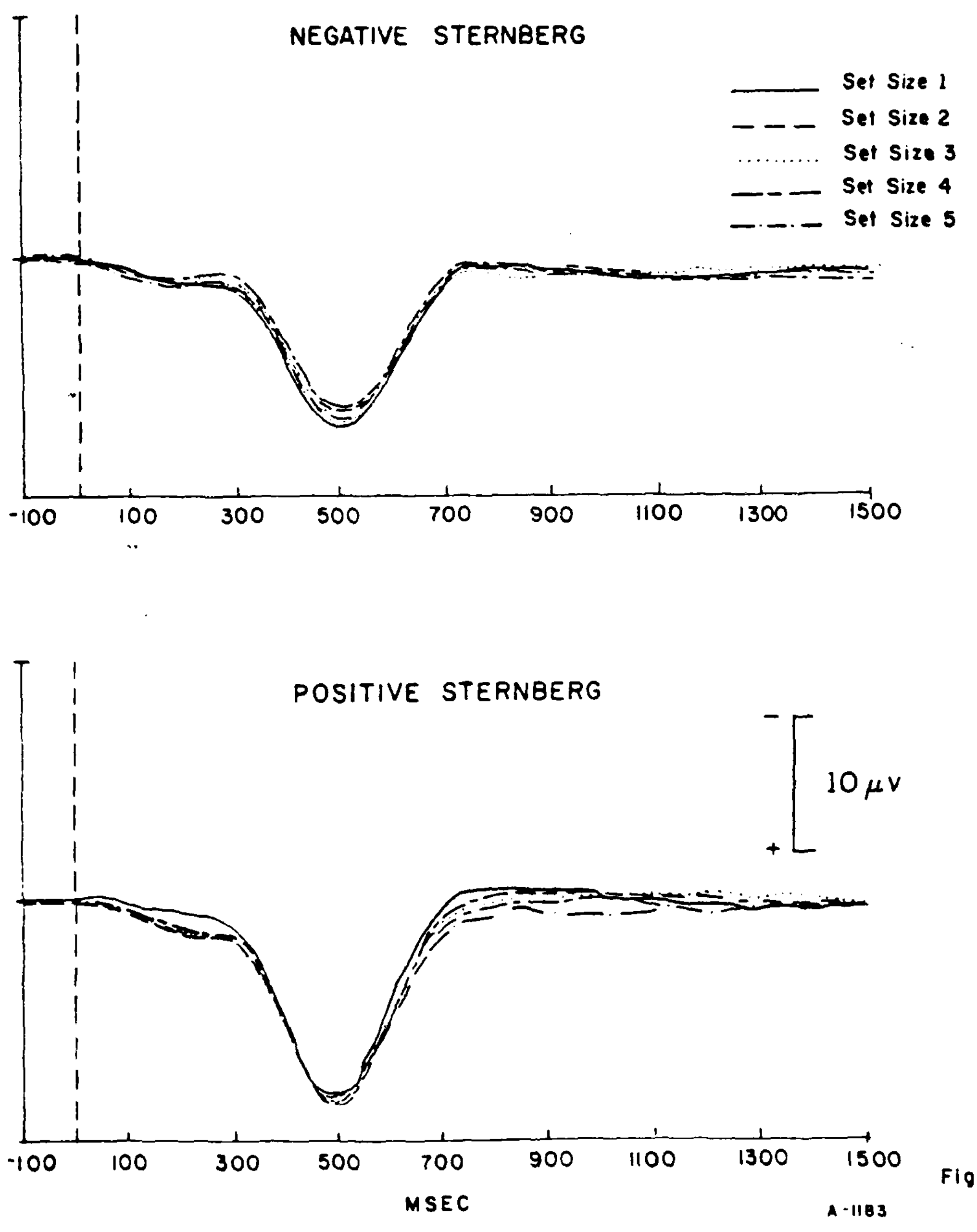

NEGATIVE STERNBERG
Set Size 1
Set Size 2
Set Size 3
Set Size 4
Set Size 5
-100
100
300
500
700
900
1100
1300
1500
POSITIVE STERNBERG
10 μv
-100
100
300
500
700
900
1100
1300
1500
MSEC

Fig. 8

A-1183

BRAIN AND INFORMATION
EVENT-RELATED POTENTIALS Vol. 425
Reprinted from
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

The Event-Related Potential as an Index of the Processing Demands of a Complex Target Acquisition Task[a]

CHRISTOPHER D. WICKENS, ARTHUR F. KRAMER AND
EMANUEL DONCHIN

Cognitive Psychophysiology Lab
University of Illinois
Champaign, Illinois 61820

Previous investigations of the effect of task difficulty on the amplitude of the P300 component of the event-related brain potential (ERP) have produced mixed results. Increases in the bandwidth of the forcing function and the number of dimensions of a compensatory tracking task increase the reaction time to an auditory probe, but have no effect on the amplitude of the P300 elicited by the probes (Wickens *et al.*, 1977; Isreal *et al.*, 1980). However, when subjects were required to monitor a simulated air traffic control display for course changes, increasing the number of elements to be monitored increased the reaction time to an auditory probe and reduced the amplitude of the P300 elicited by the probes (Isreal *et al.*, 1980). Isreal *et al.* (1980) have interpreted these results as support for a structure-specific model of processing resources (Navon and Gopher, 1979; Wickens, 1980). The investigators suggested that the amplitude of the P300 is sensitive to the perceptual demands of a task as manifested in the display monitoring paradigm while being relatively insensitive to response load, manipulated by the bandwidth and the number of axes tracked.

The effective control of a second order tracking task requires a large measure of perceptual anticipation (Wickens *et al.*, 1980). The present study was designed to test the hypothesis proposed by Isreal *et al.* (1980) by examining the effect of system order on the amplitude of the P300. A second experimental issue was the effect of practice in the primary task on the amplitude of the P300 elicited by a secondary task probe. The P300 should be sensitive to changes in the processing resource requirements of the primary task. Increased practice on the primary task should be reflected by a decrease in the discrepancy between the amplitudes of the P300s elicited during first and second order tracking conditions.

METHODS

Eight (six male) right-handed persons were recruited for Experiment 1. Eleven different (nine male) right-handed persons participated in Experiment 2. All of the subjects were undergraduate students and were paid for their participation in the study. None had any previous experience with the tracking task.

EEG was recorded from three midline sites (Fz, Cz, and Pz according to the 10–20

[a]These experiments were partially supported by a subcontract from NASA-Jet Propulsion Laboratory No. 955610 with Dr. John Hestnes as technical monitor and by the Air Force Office of Scientific Research contract F49620-79-C-0233 with Dr. Alfred Fregly as technical monitor. We gratefully acknowledge the computer programming assistance of Ron Clapman.

system: Jasper, 1958) and referred to linked mastoids. Two ground electrodes were positioned on the left side of the forehead. Burden Ag-AgCl electrodes affixed with collodion were used for scalp and mastoid recording. Beckman Bipotential electrodes, affixed with adhesive collars, were placed laterally and supra-orbitally to the right eye to record electro-oculogram (EOG) and this type of electrode was also used for ground recording. Electrode impedances did not exceed 5 kohms/cm.

The EEG and EOG were amplified with Van Gogh model 50000 amplifiers (time constant 10 sec and upper half amplitude of 35 Hz, 3dB octave roll off). Both EEG and EOG were sampled for 1280 ms, beginning 100 ms prior to stimulus onset. The data were digitized every 10 ms. ERPs were filtered off-line (−3dB at 6.29 Hz, 0 dB at 14.29 Hz) prior to statistical analysis.

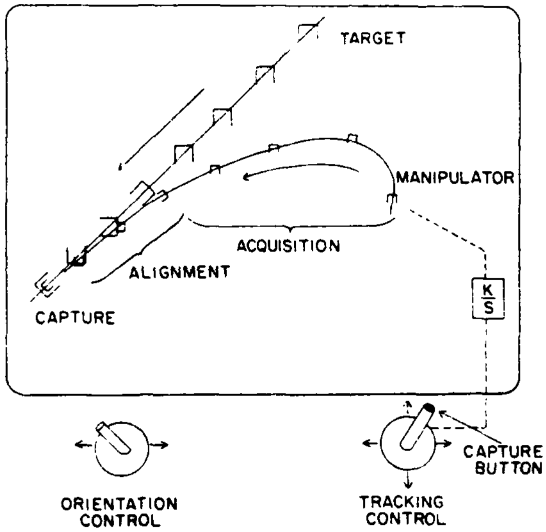


FIGURE 1. The temporal sequence of the target acquisition task (from upper right to lower left). The large three sided rectangle represents the target and the small three sided rectangle depicts the cursor. The joy stick on the right-hand side controls the path of the cursor in the x and y axes. The control stick on the left regulates the rotational velocity of the cursor.

Experiments 1 and 2 differed only in the number of practice trials the subjects received prior to the experimental blocks. The subjects in the first study performed 120 practice trials while the subjects in the second study completed 470 practice trials.

Subjects performed a three dimensional tracking task while covertly counting the total number of occurrences of a visual probe. The tracking task required the capture of a moving, rotating target using a remote manipulator system. The operator was required to match the x and y positions of the manipulator with those of the moving target and then proceed to match the angular velocity of rotation. The maximum length of a trial was 30 sec. Manipulation of task difficulty was achieved by varying: (1) the order of the control dynamics from a first order system (velocity) to a mixed first-second order (velocity and acceleration) system and (2) the experimental phase (acquisition and alignment; see FIGURE 1).

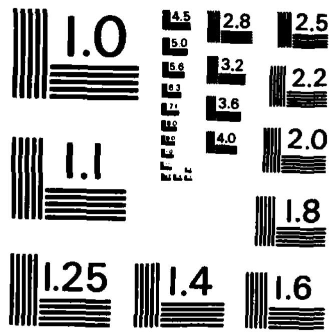
AD-A159 118 THE EVENT RELATED BRAIN POTENTIAL AS AN INDEX OF INFORMATION PROCESSING C..(U) ILLINOIS UNIV CHAMPAIGN COGNITIVE PSYCHOPHYSIOLOGY LAB E DONCHIN ET AL. 28 FEB 85 CPL-85-1 AFOSR-TR-85-0662 9/9

UNCLASSIFIED F/G 5/10 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

The experiments consisted of four blocks of thirty trials in which the subject was instructed to count intensifications of the target or cursor while performing the tracking task with either a first or second order system. Blocks were counterbalanced across subjects.

RESULTS

Counting Accuracy and Tracking Performance

There were no significant differences in accuracy of counting in the two experiments. There were differences between experiments. Subjects in Experiment 1 counted the intensifications with an average accuracy of 84% while the subjects in Experiment 2 performed at 92% ($F(1/18) = 4.98, p < 0.05$). Tracking performance was assessed in terms of the amount of time required to perform a block of trials (TT) and the number of successfully completed trials (HT). The TT was significantly longer for the second order than for the first order control dynamics in both Experiment 1 ($F(1/7) = 17.08, p < 0.01$) and Experiment 2 ($F(1/10) = 26.61, p < 0.01$). The average TT in Experiment 1 was 408 sec as compared with 281 sec in Experiment 2 ($F(1/18) = 11.7, p < 0.01$). There was a significant interaction between system order and counted symbol for HT in Experiment 1 ($F(1/7) = 16.97, p < 0.01$). The difference in HT due to control order was larger on trials when the cursor was counted. The analysis of HT in Experiment 2 indicated that HT for the second order system was larger by 15% than the HT for the first order system. ($F(1/10) = 15.66, p < 0.01$). As with the other two dependent variables, the subjects' performance as measured by HT was better in Experiment 2 (avg Exp 1 = 0.61, avg Exp 2 = 0.64).

Event-Related Potentials

The ERPs were analyzed by submitting the digitizing waveforms to a Principal Components Analysis (PCA; Donchin and Heffley, 1979). The data base for Experiment 1 was composed of 384 trials (8 subjects × 3 electrodes × 2 phases × 2 stimuli × 4 blocks) containing 128 points (1.28 sec). Experiment 2 employed eleven subjects. The Varimax rotated component scores obtained from the PCAs were analyzed in repeated measures ANOVAs.

Identification of ERP components is predicated on their latency relative to a stimulus or response, scalp distribution and sensitivity to experimental manipulations. Based on these criteria one component was identified in each of the PCAs which would qualify as the P300. Component loadings were maximal in the temporal range associated with the P300 (400–500 ms), the amplitude of the components were largest at the parietal electrode ($p < 0.01$) and the component scores were larger for the counted than the uncounted stimuli ($p < 0.01$).

Presented in FIGURE 2 are the parietal ERPs elicited by both the counted and uncounted intensifications for both phases and both experiments. FIGURE 2a (Experiment 1) shows the difference in P300 amplitude as a function of system order, instructions (counted and uncounted probes) and tracking phase. All of these effects are statistically significant ($p < 0.05$). In FIGURE 2b (Experiment 2) there are no significant differences attributable to either system order or phase. However, the effect of instruction was statistically significant ($p < 0.01$). The P300s elicited by the counted targets and cursors were not significantly different in either of the experiments.

CONCLUSIONS AND IMPLICATIONS

The results obtained in Experiment 1 demonstrated that P300 amplitude is sensitive to changes in the difficulty of a tracking task, both as a function of system order and phase. We maintain that the second order tracking demands more perceptual resources than the first order tracking because of the requirement to process higher derivatives of the error signal in obtaining stable control. The second phase of the tracking task is more difficult than the first because of the increased perceptual demands imposed upon the subjects by the requirement to control the additional rotational axis. The effect due to instruction or task relevance obtained in both of the experiments indicates that the P300 component is sensitive to subjects' allocation of processing resources. The

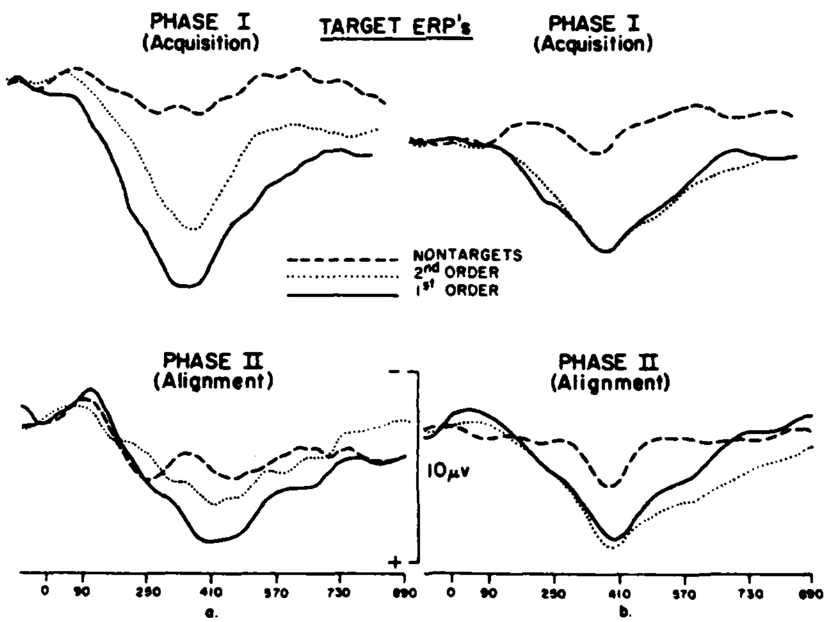


FIGURE 2. Grand average waveforms elicited by both counted and uncounted probes in Experiment 1 (a) and Experiment 2 (b).

symbol which the subjects are counting elicits a P300, while the uncounted symbol does not. The subjects' relatively high counting accuracy would support the assumption that they were performing the counting task. The differences in the waveforms across experiments can be explained in terms of the relative decrease in difficulty of the tracking task as a function of practice. Norman and Bobrow (1975) have suggested that practice tends to increase the data limited region of a task and thus free "resources" which may be employed in the processing of other tasks. This is consistent with the higher count accuracy in Experiment 2, and the absence of differences between P300s during first and second order control. Additional support for this resource modulation interpretation could be obtained by increasing the difficulty of the target acquisition task subsequent to asymptotic performance. This manipulation

should produce a difference in the amplitude of the P300 elicited in the two difficulty conditions.

The results of the present series of experiments have implications both for the study of attentional processes and the design of man-machine systems. P300 appears to provide a useful metric for determining the locus of visual attention which is not constrained by the assumption that the subject is "attending" to the area he is fixating on. This may prove to be a valuable technique in assessing subjects' allocation of attention to particular attributes of a task. In terms of the practical implications, P300 amplitude has been shown to reflect changes in both task difficulty and practice. The sensitivity of P300 to the tradeoff between these two processes may provide the instructor of operations of complex systems (*i.e.* high performance aircraft, nuclear power plant control stations) with an up-to-date model of a trainee's capability to allocate processing resources among several tasks. This information might be utilized either by the trainer or by an adaptive computer algorithm to decide when to impose additional time sharing tasks on the trainee.

REFERENCES

DONCHIN, E. & E. F. HEFFLEY. 1979. Multivariate analysis of event related potential data: A tutorial review. *In* Multidisciplinary Perspectives in Event Related Brain Potential Research. D. Otto, Ed.: U.S. Government Printing Office/EPA. Washington D.C.

ISREAL, J. B., G. L. CHESNEY, C. D. WICKENS & E. DONCHIN. 1980. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology **17**: 259–273.

ISREAL, J. B., C. D. WICKENS, G. L. CHESNEY & E. DONCHIN. 1980. The event related brain potential as an index of display monitoring workload. Human Factors **22**: 211–224.

JASPER, H. H. 1958. The ten twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. **10**: 371–375.

NAVON, D. & D. GOPHER. 1979. On the economy of the human processing system. Psychol. Rev. **86**: 214–255.

NORMAN, D. & D. BOBROW. 1975. On data limited and resource limited processes. Cognitive Psychology **7**: 44–64.

WICKENS, C. D. 1980. The structure of attentional resources. *In* Attention and Performance VIII. R. Nickerson & R. Pew, Eds.: Erlbaum. Hillsdale, N.J.

WICKENS, C. D., J. B. ISREAL & E. DONCHIN. 1977. The event related cortical potential as an index of task workload. *In* Proceedings of the Human Factors Society, 21st Annual Meeting. A. S. Neal & R. F. Palasek, Eds.: Santa Monica, Ca.

WICKENS, C. D., W. D. DERRICK, J. MICALLIZI & D. BERRINGER. 1980. The structure of processing resources. *In* Proceedings of the Human Factors Society, 24th Annual Meeting. R. E. Corrick, E. C. Haseltine & R. T. Durst, Eds.: Los Angeles.

END

FILMED

11-85

DTIC